Information Asymmetries and Incentives in Health Care Markets

Dissertation
for the Faculty of Economics, Business Administration
and Information Technology of the University of Zurich

to achieve the title of
Doctor of Philosophy
in Economics

presented by
Maria Trottmann
from Baden / Rottenschwil (AG)

approved in October 2011 at the request of
Prof. Dr. Peter Zweifel
Prof. Dr. Michelle Goeree

The Faculty of Economics, Business Administration and Information Technology of the University of Zurich hereby authorises the printing of this Doctoral Thesis, without thereby giving any opinion on the views contained therein.

Zurich, 26. October 2011

Chairman of the Doctoral Committee: Prof. Dr. Dieter Pfaff

Preface

This doctoral thesis was written during my time at the University of Zurich, the CSS Institute for Empirical Economics, and Boston University. I was supported by many excellent people to whom I wish to express my gratitude.

First and foremost, I would like to thank my supervisor Prof. Dr. Peter Zweifel for his invaluable support and intellectual input. In addition to sharing his deep knowledge of health and insurance economics, he created an inspiring work environment, which he was generous enough to let me use long before I became a member of his team. Second, I am indebted to Prof. Dr. Konstantin Beck, who motivated me to write this thesis while I was a member of his team at CSS insurance. Prof. Dr. Beck also provided data, intellectual input and much-valued feedback. Third, I enjoyed the privilege of visiting Prof. Dr. Randall Ellis at Boston University for six months in 2009. I would like to thank Prof. Dr. Ellis for sharing his precious time and in-depth understanding of health economics. Fourth, I wish to thank Prof. Dr. Michelle Goeree, who not only organized many inspiring seminars in applied microeconomics, but has also agreed to grade this thesis.

Additionally, I would like to thank my former and current colleagues at the University of Zurich, Maurus Rischatsch, Philipp Morf, Dr. Ilja Neustadt, Dr. Johannes Schoder, Dr. Michèle Sennhauser, Dr. Harry Telser, Phillipe Widmer, Adrian Rohner, Dr. Patrick Eugster, Dr. Boris Krey, Micha Ruflin, and PD Dr. Felix Schläpfer, for inspiring discussions, programming help, constructive feedback, coffee breaks, etc. I also want to thank the whole team at the library for help with my literature requests,

and the team at the dean's office for help with the administration. Special thanks goes to Philipp Morf and Crawford Smith for excellent proof reading.

I would also like to thank my former colleagues at the CSS insurance company, especially Monika Buholzer, Urs Kaeser, Bernhard Keller, Ute Kunze, Ute Studer, Stefan von Rotz, and Dr. Longin Korner. During my time at CSS, I experienced the daily business of providing health insurance in a strongly regulated market. This experience was of high value during the writing of this thesis. In addition, I want to thank my former colleagues at Verisk Health in Waltham, MA (USA), and Munich (Germany). Working with experts on risk adjustment schemes in the US and Germany was a privilege. I also wish to thank the members of the European risk adjustment network RAN. The yearly RAN conferences were illuminative and motivating.

This is also to thank my parents Beatrix and Joerg Trottmann, whose continuing support enabled me to do such a long education, and my grandfather Theo Ortner, whose never-fading enthusiasm for this thesis motivated me during many phone calls and personal meetings.

Most of all, I want to thank my partner in love and life, Sven Ruef, for all his support and patience during these years.

THANK YOU VERY MUCH!!

Maria Trottmann
Zurich, June 2011

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This collection of essays explores issues related to asymmetric information in health care markets. Prominent topics are (1) the structural characteristics that have emerged in health care markets in response to information asymmetries, (2) the incentives of market players and, (3) their responses to these incentives.

A first salient feature in health care markets is that - in addition to the usual sellers (physicians) and buyers (patients) - important players include third-party payers (insurers) and government agencies (regulators). The case for insurance is undisputed because individual demand for health care services is highly unpredictable. Indeed, an individual might wish to spend nothing on health care on one day and an amount exceeding her total wealth on the next, an incidence that is not observed for other commodities [Newhouse (2002)]. The case for government involvement is made for several reasons, which are listed by Zweifel et al. (2009). Three of them will be briefly mentioned here. First, patients might be unable to make rational decisions regarding their consumption of medical care because of their insufficient information about treatment possibilities. Second, private markets for health insurance might fail to provide the desired amount of coverage to all citizens due to problems of adverse selection and premium risk (denoting the risk that individuals face premium hikes after falling ill). Third, medical care is a good with high 'altruism', meaning that

many members of society wish to make medical care available to all their fellow citizens.

Information asymmetries occur between all players in health care markets. Three examples are analyzed in this dissertation. In Chapter 2, the regulator wishes to offset incentives for risk selection, but has an information disadvantage vis-à-vis health insurers. In Chapter 3, health insurers are confronted with moral hazard because they cannot fully observe the actions of patients and physicians. In Chapters 4 and 5, physicians are better informed about optimal treatment paths than patients and therefore take decisions on their behalf (physician agency).

Chapter 2 analyzes a market with 'managed competition' between health insurers [Van de Ven et al. (2007)]. The regulator wishes to offset insurers' incentives to select good risks by paying out risk-adjusted premium subsidies. However, the regulator can neither perfectly observe an insurer's risk pool nor does he know the cost of offering coverage efficiently. Nevertheless, calculations in Chapter 2 show how risk adjustment schemes can be designed in a way to substantially reduce incentives for risk selection, even with imperfect information. Moreover, the proposed schemes can turn elderly or even chronically ill individuals into favorable risks from the insurers' point of view. In that way, health-based risk adjustment might spur investments away from risk selection activities towards efficiency improvements [Van de Ven and Ellis (2000)]. Examples of such investments are contracts with increased cost sharing, which are analyzed in Chapter 3.

Chapter 3 contains an empirical estimation of how demand-side and supply-side cost sharing can be used as a corrective for moral hazard. Moral hazard occurs because health insurers can neither observe their clients' efforts to prevent losses ex-ante, nor verify the size of the loss ex-post. Therefore, they base payments on the consumption of medical care. If patients are fully insured and providers are paid fee-for-service, the desired quantities of medical care exceed the efficient level. Both demand-side cost sharing plans (in the guise of higher deductibles) and supply-side cost sharing plans (in the guise of capitated independent practice associations (IPA)) are offered by

Swiss health insurers in exchange for a premium rebate. It is found that both types of cost sharing can reduce health care expenditure. However, when the trade-off between moral hazard reduction and risk selection effects is considered, supply-side cost sharing is somewhat more effective. One reason why supply-side cost sharing is applied is the fact that physicians are better informed about the possibilities and consequences of treatment, and therefore they are delegated a great deal of decision making authority. This agency relationship is analyzed in more detail in Chapters 4 and 5.

Chapters 4 and 5 analyze the impact of financial incentives on physician decision making and physician agency using the fact that some Swiss physicians have the right to dispense drugs on their own account, while other physicians prescribe drugs only. The former group has financial interests attached to the type and quantities of drugs sold, while the latter group does not.

In Chapter 4, a prescribing physician's dispensing status is related to her choice of generic versus brand-name drugs. It is found that physicians who dispense on their own account are more likely to prescribe generic drugs compared to physicians who do not. This is likely due to higher contributions to income from generic drugs. While this points to imperfect agency, no distinct evidence was found for the expectation that physician agency for the patients is reduced if she dispenses on her own account compared to a situation where she does not. Physicians in managed-care type arrangements are more likely to prescribe generic drugs than other physicians, which might well be one of the reasons for the reduced expenditure observed in the IPA plan in Chapter 3.

In Chapter 5, a general practitioner's (GP) dispensing status is related to her incentives (1) to prescribe drugs, (2) to provide treatments, or (3) to make referrals to hospitals. From a theoretical model, it is expected that a dispensing GP would prescribe more drugs, but provide less GP services and make less referrals than a non-dispensing GP. However, in the empirical estimation, only the last hypothesis could be confirmed. In particular, it is found that patients who buy drugs from

their physicians have slightly lower drug expenditure than patients who buy at the pharmacy, which is in line with their increased use of generic drugs that was found by in Chapter 4. Referral rates to hospitals were found to be lower among dispensing physicians, maybe due to their higher incentives to keep their patients.

These four essays discuss examples where - in spite of considerable information asymmetries - market mechanisms might be useful to achieve society's goals in health care markets. Results in Chapter 2 imply that health-based risk adjustment can be designed in a way to ensure equal access for high risk in competitive health insurance markets, while at the same time increasing incentives for investments in efficiency improvements. The demand- and supply-side cost sharing options discussed in Chapter 3 attenuate moral hazard while being voluntary chosen by insured individuals in exchange for a premium rebate. In Chapter 4, it was found that the financial incentives of dispensing physicians might foster rather than undermine generic substitution. Last but not least, results in Chapter 5 imply that physicians with dispensing rights might reduce referral rates to hospitals in response to financial incentives.

Note that Prof. Dr. Peter Zweifel co-authored Chapters 2, 3 and 4, Prof. Dr. Konstantin Beck co-authored Chapters 2 and 3, and Maurus Rischatsch co-authored Chapter 4. Chapter 2 is published in the Journal of Health Economics. Chapter 3 has been accepted for publication in the Journal of Health Economics. Chapter 4 has been submitted to the International Journal of Health Care Finance and Economics, and Chapter 5 has been submitted to the European Journal of Health Economics.

This introduction is concluded with a note concerning the structure of this dissertation. Each chapter of this dissertation can be considered as self-contained, having its own appendix. Institutional features are explained separately in each chapter in the interest of readability. References across chapters are made explicit.

Risk Adjustment in Health Insurance and its Long-term Effectiveness

Konstantin Beck, Maria Trottmann and Peter Zweifel

Abstract: This paper seeks to create new insights when judging the impact different risk adjustment schemes may have on the incentive to select risks. It distinguishes risk types with high and low profit potential and estimates long-run profits associated with risk selection in four scenarios (no risk adjustment, demographic only, including prior hospitalization, and including prior hospitalization and Pharmaceutical Cost Groups). The database covers 180,000 Swiss individuals over 8 years, 3 of which are used for model building and 5, to estimate insurers' profits due to risk selection in the four scenarios. While these profits prove to be very high without risk adjustment and still substantial with demographic risk adjustment, they become surprisingly low when the crude morbidity indicator 'prior hospitalization' is included in the formula. These results clearly indicate the need for health status-related risk adjustment in insurance markets with community rating, taking into account insurers' planning horizon.

Keywords: Risk adjustment, Risk selection, Managed competition
JEL classification: G22; I10; I11

Published in 'Journal of Health Economics (29) 2010, p. 489 - 498'.

Chapter 2

Risk Adjustment in Health Insurance and its Long-term Effectiveness

2.1 Introduction

Enthoven's proposal for regulated competition between social health insurers [Enthoven (1978)] has been used as a blueprint for reform in several countries [see Van de Ven et al. (2007)]. One example is Switzerland with its comprehensive mandatory coverage for all citizens, offered by some 80 competing nonprofit health insurers. The new law of 1994 calls for semiannual open enrollment and community-rated premiums within the same fund. Premium reductions for adults within a given fund are only possible for contractual differences such as a higher deductible. However, with every adult paying the same premium - within a given fund for the same type of contract - but expected health care expenditure (HCE) varying widely, strong incentives for risk selection are created in the absence of an adequate risk adjustment scheme. Although risk selection is illegal, its prevalence in Swiss social health insurance has been reported repeatedly [Beck and Zweifel (1998), Beck et al. (2003)]. As Van de Ven and Ellis (2000) argue, risk selection produces no benefits to society (unless a dynamic view is adopted, where the threat of being classified as an unfavorable risk in the future helps to reduce moral hazard).

The objective of risk adjustment is to mitigate incentives for risk selection. To this end, insurers with a below-average share of female and elderly consumers have to contribute to the risk adjustment fund, while insurers with an above-average share receive a payment from the fund. So far, the different schemes have been judged mainly in terms of their ability to predict individual HCE one year ahead [Newhouse et al. (1989), Van Barneveld et al. (2000), Holly et al. (2003)].

This criterion is subject to at least two criticisms. First, risk selection is not costless to insurers. As pointed out by Van Barneveld et al. (2000) as well as Zweifel and Breuer (2006), this means that they will invest in this activity only if expected profits exceed the cost. However, the regression criterion of minimizing squared prediction error with regard to HCE fails to take cost considerations into account. In our model, we address the problem of costly risk selection activities by restricting attention to selection profits and losses exceeding a given annual threshold. Second, Zweifel and Breuer (2006) argue that insurers who want to stay in business must have an eye on present values rather than one-period profits.

This paper, then, follows the lead of Shen and Ellis (2002) by estimating expected profits attainable from risk selection, given the information available to the insurer. It therefore only models the classification of risk types, neglecting the problem of how to attract or deter types. However, if profits are large enough, strategies to perform risk selection will most certainly be developed by crafty insurers.

However, unlike Shen and Ellis (2002), the present analysis assesses the impact of risk adjustment if insurers' planning horizon exceeds one year. In an attempt to reflect longer planning horizons (which agree with insurers' preference for long-run contracts and guaranteed renewability unless prevented by regulation [Pauly et al. (1998)]), the period of observation for expected profits is extended to five years in the body of the paper. This permits to take into account the fact that a currently favorable risk may develop into an unfavorable one, switch to a competitor, or die. Conversely, an unfavorable risk may recover to become a favorable one in the future. In the theory of statistics, these effects are known as 'regression towards the mean'

[Welch (1985); Beck (2004)]. The empirical relevance of the regression towards the mean effect is assessed by varying the planning horizon from one to five years. This is possible thanks to a panel data set covering some 180,000 individuals over eight years.

The reminder of this article is structured as follows. In Section 2.2, a model of the insurer's decision to select risks is formulated, which subsequently permits to calculate the financial reward from this activity. After a description of the risk adjustment schemes and the database in Section 2.3, the details of the empirical estimation are explained in Section 2.4. Results are presented in Section 2.5. They indicate that even a crude adjustment in the risk adjustment formula to take future HCE into account serves to neutralize incentives for risk selection to a substantial extent also over a longer planning horizon. The final Section 2.6 is devoted to a summary and conclusions.

2.2 Modeling Risk Selection

2.2.1 What Risks to Attract or Deter?

The objective of this section is to model a health insurer's decision to attract or deter certain risk types. This decision is assumed to reflect insurers' expected profits or losses ($E[\pi_{i,j}]$) pertaining to customer (i), taking into account the risk adjustment formula (j) applied by the regulator. To estimate this quantity, five elements must be considered; viz. (1) the expected fair community-rated premium ($E[P_{i,t,j}]$), (2) expected health care expenditure ($E[HCE_{i,t}]$), (3) the expected contribution to the risk adjustment scheme ($E[RA_{i,t,j}]$), a positive quantity for favorable risks, a negative one for unfavorable ones, its value depending on the risk adjustment formula (j),[1] and (4) the probability of an individual dying ($p_{i,h}^{death}$) or (5) switching to a competitor

[1] All formulas are constructed in a way to guarantee that the sum of contributions to risk adjustment paid by low risks is exactly balanced out by the sum of subsidies paid out for high risks.

($p_{i,k}^{switch}$). This all boils down to Expression (2.1), with the interest rate r (set to 0.06 throughout this paper) used to discount future payments,

$$E[\pi_{i,j}] = \sum_{t=2000}^{2004} (E[P_{i,t,j}] - E[HCE_{i,t}] + E[RA_{i,t,j}]) \prod_{h=2000}^{t} (1 - p_{i,h}^{death}) \prod_{k=2001}^{t} (1 - p_{i,k}^{switch}) \frac{1}{(1+r)^{(t-2000)}}. \quad (2.1)$$

Loadings for administrative expenses are neglected, because they are part of ($E[P_{i,t,j}]$) and ($E[HCE_{i,t}]$) but do not enter ($E[RA_{i,t,j}]$). Note that HCE does not depend on the type of risk adjustment imposed, although insurers' incentives to combat moral hazard (by launching product innovations) may well be weakened by risk adjustment [Zweifel (2007)]. The assumed planning horizon comprises the years 2000 (defined as 'current') to 2004. Insurers' actual planning horizons might be even longer; however, data availability dictates one of no more than five years. This should be sufficient to at least approximate expected long-term profits. Following the approach proposed by Van Barneveld et al. (2000), only 'sufficiently large' profits or losses are assumed to cause risk selection activities. Profits in principle are 'sufficiently large' if returns to risk selection exceed its total cost, which not only comprises the expenses for product development, marketing, and actual administration of the risk portfolio but also the loss of reputation if found out by the media or the regulator. Clearly, information to estimate this quantity is not publicly accessible. Therefore, it is simply assumed that expected profits from risk selection must exceed CHF 1,000 (= $800 at 2006 exchange rates) per annum and individual in present value to trigger selection activities. In a sensitivity analysis, results changed surprisingly little when the threshold was lowered to CHF 400 and increased to CHF 1,200 p.a. The impact of this ad hoc assumption is therefore limited.

According to Table 2.1, all customers are divided into four mutually exclusive subsets. Group A contains all individuals with expected profits in excess of CHF

Table 2.1: Assumed Risk Selection Strategies, 1 CHF $\approx$ 0.8\$.

Risk type	Characterization	Strategy
A	Expected profit $>$ 1,000 CHF p.a.	Attract
B	Expected profit $\leq$1,000 CHF p.a.	Passive
C	Expected loss $\leq$1,000 CHF p.a.	Passive
D	Expected loss $>$ 1,000 CHF p.a.	Deter

1000 p.a., while B has those with expected profits up to CHF 1,000 p.a. Conversely, C contains all individuals with expected losses up to CHF 1,000 p.a. and D those with losses beyond CHF 1,000 p.a. Therefore, A is the set of risks the insurer seeks to attract, D contains the risks it wants to deter, while B and C are the risks that do not call for any risk selection effort. It is important to note that risk selection does not describe a 'young-and-healthy-people-only' strategy under all circumstances. As shown in Section 2.5.2, risk adjustment can turn elderly and even chronically ill individuals into favorable risks.

2.2.2 Profits Due to Risk Selection

In order to assess the effectiveness of risk adjustment, the insurer's profits are calculated under the assumption that it had successfully applied the selection strategy described by Table 2.1. The ex-post or realized profits generated by individual i and associated with risk adjustment scheme j are specified as follows,

$$\pi_{i,j} = \sum_{t=2000}^{2004} (P_{i,t,j} - HCE_{i,t} + RA_{i,t,j}) \frac{1}{(1+r)^{t-2000}} \frac{1}{\tau} \tag{2.2}$$

with

$$\tau_t = \frac{\sum_i HCE_{i,t} / \sum_i m_{i,t}}{\sum_i HCE_{i,2000} / \sum m_{i,2000}}.$$

Here τ_t is a deflator and $m_{i,t}$ is the number of months individual i is enrolled during a given year. Since τ_t reflects the general development of HCE since 2000, real profits are defined in terms of the cost of health care. Transformation to real terms facilitates the comparison of ex-post profits with ex-ante expectations. The latter [see Equation (2.1)] are calculated in real terms for simplicity since risk classification is not affected by general inflation. Considering first the strategy of deterring unfavorable risks, the financial benefit attainable, Π, is defined as profits contributed by the remaining risk types (i.e. A, B, and C) relative to their total (deflated) premium revenue:

$$\Pi_{[deterD]} = \frac{\sum_{i\epsilon\{A,B,C\}} \pi_{i,j}}{\sum_{i\epsilon\{A,B,C\}} \sum_{t=2000}^{2004} P_{i,t,j}(1/\tau_t)(1/(1+r)^{(t-2000)})}. \tag{2.3}$$

Note that while labels A, B, C denote those individuals the insurer expects to be preferred or neutral risks (according to Table 2.1), Formula (2.3) calculates realized profits from A, B and C-type customers. The same is true for Formula (2.4) below. Realized figures serve to simulate the future profits associated with the choice of a risk-deterring strategy in the year 2000.

The other strategy, attracting good risks, has to be defined differently because it is inconceivable that the population insured consist of subset A only. Rather, let $x > 1$ be the factor by which the size of A is increased (for example by strategically promoting high-deductible plans known in social health insurance in the Netherlands or Switzerland to provoke self-selection by low risks). Then, realized financial benefit is given by:

$$\Pi_{[attractA]}$$

$$= \frac{(x-1)\sum_{i\epsilon A} \pi_{i,j}}{(x-1)\sum_{i\epsilon A} \sum_{t=2000}^{2004} P_{i,t,j}(1/\tau_t)(1/(1+r)^{t-2000}) + \sum_i \sum_{t=2000}^{2004} P_{i,t,j}(1/\tau_t)(1/(1+r)^{t-2000})}. \tag{2.4}$$

If, for example, an insurer is able to triple its share of preferred A-type customers, its realized profits will increase by $2\sum_{i\epsilon A} \pi_{i,j}$ relative to a pre-

mium volume that itself increases by the first term in the denominator, i.e. $2 \sum_{i \epsilon A} \sum_{t=2000}^{2004} P_{i,t,j} \ (1/\tau_t)(1/(1 + r)^{t-2000})$. The more effective the risk adjustment, the smaller are these profits. As Beck and Zweifel (1998) point out, however, risk selection is a risky business. Some customers who are expected to be profitable will in effect turn out to be unfavorable risks, while some who are deemed unfavorable will in fact contribute to profits. As will be seen in Section 2.5.3, this uncertainty increases when the risk adjustment formula is refined.

2.3 Policy Setting and Data

2.3.1 Risk Adjustment Schemes Considered

The profits of risk selection are assessed in four different scenarios. The $RA_{i,t,j}$ values appearing in Equations (2.1) and (2.2) are calculated according to the four risk adjustment schemes. They are retrospective rather than prospective, as in current Swiss regulation.

(0) No risk adjustment
The first scenario is a benchmark with no risk adjustment scheme in place.

(1) Demographic risk adjustment
Current Swiss risk adjustment uses 28 age and gender groups (26-30, 31-35 ... 91+), as in Table 2.4. While these age groups are purely arbitrary, they are established in the market. The possibility of optimizing them for risk adjustment will therefore not be considered here. Also, payments are calculated regionally. To avoid small sample problems in the top age groups of small cantons, this detail is neglected by treating Switzerland as one region.

(2) Demographic risk adjustment augmented by prior hospitalization
This formula is part of a reform passed by Swiss parliament in 2007. Retaining cur-

rent age and gender groups, it includes a dummy variable indicating hospitalization during the previous year. Empirical evidence presented by Beck (1998) and Holly et al. (2003) shows HCE to be substantially higher for individuals with hospital stays[2] during the previous year. However, insurers might have an incentive to encourage short-term hospitalizations with the mere aim of receiving payments from the risk adjustment scheme. Therefore, only inpatient stays of three or more days are considered to be a hospitalization. Note that it is not the cost of the inpatient stay itself that is taken into account but the increased predicted HCE during the year following the stay. Therefore any manipulation of this adjuster would pay off only if this extra HCE were to exceed the cost of the hospitalization itself, which is very unlikely in the case of long stays.

(3) Demographic risk adjustment augmented by prior hospitalization and PCGs
The final alternative to be considered is to augment existing demographic risk adjustment by both the indicator for prior hospitalization and Pharmaceutical Cost Groups (PCGs). There are 13 PCGs which are similar to those developed by Lamers and Van Vliet (2003). They were adapted to Swiss data by a team at CSS [Beck et al. (2006)].

As with all patient classification systems, the issue of how to deal with patients belonging to more than one class has to be addressed. Here, the sorting algorithm used by Pope et al. (2000) is employed by first calculating mean HCE by PCG for the entire sample and assigning the PCG with the highest value rank one and excluding its members from further calculations. Next, mean HCE is recalculated for the reduced sample, assigning the PCG with the highest value rank two, and so on. Finally, patients with more than one condition are assigned to the PCG with the highest rank (Table 2.2).

[2]Inpatient stays related to maternity are excluded.

Table 2.2: Ranking of Pharmaceutical Cost Groups in Terms of HCE.

Rank	PCG	Population share, in percent	Average excess HCE, in CHF per month
1	Renal disease, ESRD	0.06	3,484
2	HIV / AIDS	0.11	1,529
3	Transplantations	0.15	1,291
4	Malignancies	0.13	970
5	Diabetes, insulin-dependent	0.75	558
6	Morbus Parkinson	0.38	440
7	Epilepsy	0.89	280
8	Respiratory illness & Asthma	2.16	248
9	Morbus Crohn, Colitis ulcerosa	0.23	215
10	Diabetes, non insulin-dependent	2.4	180
11	Rheumatologic conditions	2.85	165
12	Acid peptic disease	0.59	142
13	Cardiac disease	3.96	114
0	None	85.33	-

1 CHF $\approx$ 0.8 US$

The predictive power of the four risk adjustment formulas is shown in Table 2.3. The R^2 values are high, mainly for two reasons. First, Swiss health insurers only pay roughly one-half of inpatient HCE, in keeping with the Law of Health Insurance of 1994. One half of the bill is funded by cantonal governments, who heavily subsidize public hospitals. Since very high HCE are almost always due to hospitalization, outliers do not fully show in the data, causing goodness of fit to increase. Second, the marked increase in R^2 from variant (2) to variant (3) can be explained by the fact that little prediction is involved because the observations on PCGs pertain to the same year as those on HCE. Table 2.3 also shows that even with PCG information included, the regulator (who has to disregard prior HCE for maintaining health insurers' incentive to control cost) cannot catch up with the health insurers, who dispose of a good deal of additional information which can be used to increase R^2 to 0.48 (see Section 2.4.2 for details).

Table 2.3: The Four Risk Adjustment Formulas Compared.

Risk adjustment formula No.	R^2, Year: 2000
(0) None	-
(1) Age, gender	0.11
(2) Age, gender, prior hospitalization	0.21
(3) Age, gender, prior hospitalization, PCG	0.30
Benchmark: Insurer's own model	0.48

2.3.2 The Data

The sample contains individual data from 182,529 adults (aged 26+) enrolled by CSS, the leading sickness fund of Switzerland, during the full year of 1999 and not enrolled in a Managed Care plan during the period of observation. For data quality reasons, only residents of the French and Italian speaking parts of Switzerland are included.[3] Individuals were observed from 1997 to 2004, with 1997 to 1999 used for prediction. The insurer is assumed to undertake its risk selection effort once and for all at the beginning of the year 2000. The data from 2000 to 2004 serve for calculating the present value of profits it would have made by pursuing the respective strategy.

2.4 Calculating the Components of Profits and Losses

To calculate expected profits from risk selection according to expression (1), all components such as expected premiums, expected HCE, expected payments into and from the risk adjustment fund as well as probabilities of death and of switching to a competitor need to be determined. This section is devoted to these issues.

[3]Due to different billing modalities, detailed information on drug expenditure is of good quality only in the French and Italian speaking parts of the country. Except for the Pharmaceutical Cost Group variant of risk adjustment, the analysis presented in Sections 2.4 and 2.5 was repeated using a larger sample containing 250,000 insured from all parts of the country. Results were very similar to those presented here.

2.4.1 Expected Premiums

In a competitive market with entry and exit, total premium revenue equals total expected cost, the latter made up of expected HCE plus a loading. Since the admissible loading (of about 5 percent) is part of premium revenue as well as of cost, it does not affect individual contributions to profit and is therefore neglected.[4] Premiums must be community-rated for all adults within the same fund, region, and coverage option (e.g. deductible level). Premium reductions for high-deductible plans are possible, but - due to regulation - fall short of their risk-rated amounts. Contracts with high deductibles are especially attractive to low risks, causing them to be an effective means for risk selection. Perfect risk adjustment would neutralize these incentives; however, given imperfect adjustment formulas - and all formulas analyzed in this study are imperfect - differences in expected profits across deductible options remain. As a consequence, high-deductible plans cross-subsidize low-deductible plans. Therefore consumers choosing high-deductible plans become preferred risks to insurers.

In order to be able to use observed values, the insurer considered is assumed to have predicted total HCE of the benchmark year 2000 with perfect precision. Moreover, to simplify calculations, inflation during the forecasting period 2001 - 2004 is neglected. In fact, as long as inflation affects all components of Equation (2.1) in the same way (including payments to/from the risk adjustment scheme), real profits do not change. Therefore, calculated premiums $E[P_{i,t,j}]$ are constant over these 4 years.

2.4.2 Expected Health Care Expenditure

Predicted individual HCE is derived from prior experience, covering the years 1997 - 1999. The year 1999 was complete, while missing entries from 1997 to 1998 were replaced by the average values pertaining to their demographic group. Insurers know past individual HCE for existing enrollees. For new enrollees, they can predict HCE

[4] A detailed analysis of administrative expenses might show different loadings for different risk groups; however, this goes beyond the scope of this study.

using information from questionnaires that have to be filled by new applicants for supplementary insurance (which is regulated differently, by the Law on Insurance Contracts). Only switchers having no more than compulsory coverage are not made to declare their health status. Still, sales personnel obtains an (often visual) impression of the customer's health. The choice of deductible for the compulsory part also helps to predict HCE.

Future individual HCE is estimated in three steps. The first is an OLS regression with HCE in 2000 net of deductible and coinsurance as the dependent variable. Despite the fact that about 30 percent of individuals did not consume health care in excess of their deductible and the high skewness of positive HCE, untransformed OLS estimation is the preferred method of estimation. In a comparison of alternative models (e.g. a two-part logarithmic model with 'smearing estimate' retransformation [Duan et al. (1982)] and a GLM model with a log link and a gamma family [Manning and Mullahy (2001)]), untransformed OLS was found to have the smallest mean squared and smallest mean absolute prediction error. It performed particularly well at predicting very high HCE, which is of crucial importance to the insurer. Similar findings have been reported by e.g. Pauly and Herring (1999), Holly et al. (2003), and Ellis (2008). The explanatory variables are age classes interacted with gender, deductibles as of year 2000, HCE in 1997, 1998, and 1999 (the latter split up in its components, viz. physician's services, drugs dispensed by physicians, drugs dispensed by pharmacies, inpatient care, home care, nursing home care, and other expenditure).

The regression results appear in Table 2.4, with its first three columns showing age and gender effects, while its last two columns contain the estimates pertaining to the remaining regressors. Since the normality assumption does not hold in view of skewness, distribution-free Tchebycheff significance levels are also reported.

Except for the constant which is negative, all coefficients have the expected signs. For women, HCE attains a maximum (ceteris paribus) in the 26-30 age group (due to maternity), but rises consistently between age groups 41-45 and 86-90. Beyond age 90 (men 85) age coefficients go down. Otherwise, men display a consistent increase

Table 2.4: Individual HCE net of Deductible and Coinsurance, 2000.

Age groups	Female	Male	Other Factors	
			Constant	-354**°
26 - 30	492**°	reference cat.	HCE 97	0.109**°
31 - 35	386**°	28	HCE 98	0.080**°
36 - 40	229*	102	Physician services 99	0.540**°
41 - 45	181*	122	Drugs doctors 99	0.947**°
46 - 50	227*	229*	Drugs pharmacies 99	0.977**°
51 - 55	321**°	275**	Inpatient care 99	0.347**°
56 - 60	310**°	424**°	Home care 99	0.936**°
61 - 65	472**°	592**°	Nursing home care 99	0.626**°
66 - 70	643**°	1,042**°	Other HCE 99	0.626**°
71 - 75	1,108**°	1,307**°	Deductible 230	589**°
76 - 80	1,602**°	1,725**°	Deductible 400	297**°
81 - 85	2,156**°	2,072**°	Deductible 600	86
86 - 90	2,886**°	2,666**°	Deductible 1,200	79
91+	2,580**°	2,267**°	Deductible 1,500	reference cat.

$R^2 = 0.481$, $R^2_{adj} = 0.481$, F(df=40) = 3,750**, n = 182,529, *p $\leq$ 0.05 , ** p $\leq$ 0.01, °Tchebycheff significance level 10%

of HCE with age. In a second step, individual HCE in 2004 is predicted using the 1997-1999 values of explanatory variables (age as of 2004). Negative predicted values (occurring among about 5 percent of insured) are replaced by zeroes. The third step consists in interpolating between the predicted 2004 and the observed 1999 values, in accordance with Equation (2.5). This procedure can be justified by noting that observed HCE contains transitory components, while the predicted 2004 values are purged of them. By smoothing HCE values prior to 2004 as well, interpolation serves the accuracy of prediction. For the choice of the interpolation formula, a natural assumption is that insurers increasingly discount the transitory component as time goes by. Using exponential decay, the formula reads:

$$E[HCE_{i,t}] = E[HCE_{i,2004}] - (0.05)^{t-1999} E[HCE_{i,2004}] - E[HCE_{i,1999}] \qquad (2.5)$$

$$\text{for} \quad t \ \epsilon \ \{2000, 2001, ..., 2003\}$$

It amounts to moving average variant (MA) of Van Vliet (1992) and his implementation of the 'regression to the mean hypothesis' by Welch (1985). All three variants (a simple autoregressive (AR) model, an AR variance component model, and an ARMA model) were fitted to the residuals of a HCE regression using individual data of 33,987 Swiss insured from 1990 to 1997. The three specifications yielded very similar AR coefficients, ranging between 0.491 and 0.521 [Beck (2004)]. These findings support the use of Formula (2.5).

2.4.3 Expected Payments from / to the Risk Adjustment Fund

The ex-post calculation of the different risk adjustment formulas is straightforward. By way of contrast, modeling the insurer's expectations ($E[RA_{i,t,j}]$ in expression (2.1)) raises a few issues. When applying the different risk adjustment schemes to future years (2001–2004), age is known while gender and existing chronic illness can be assumed to remain constant. Individuals who will develop chronic conditions are aggregated with the low risks, assuming that they cannot be identified. By way of contrast, knowing that about 12 percent of enrollees had at least one hospital stay during the preceding year, the insurer associates by assumption the top 12 percent in terms of total HCE with those that will have a hospital stay (relevant for risk adjustment Schemes (2) and (3)).

2.4.4 Probability of Leaving the Insurer

As evidenced by Equation (2.1), two probabilities need to be distinguished here, the probability of death, ($p_{i,h}^{death}$) and the probability of switching to another insurer, ($p_{i,k}^{switch}$). Estimating ($p_{i,h}^{death}$) specifically for this population turned out to be impossible because of the small number of deceased in the sample. Instead, life tables provided by the Federal Statistical Office were used, which are grouped by age and gender. However, high HCE have been found to be strongly related to death by e.g. Zweifel et al. (2004). The probability of death is therefore certainly underrated for

high-cost individuals, since they are more likely than others to drop out of the sample in the course of the forecasting period. This results in an overestimate of expected HCE.

To estimate the probability of an insured switching to a competitor ($p_{i,k}^{switch}$), a logistic regression model is used [Beck (2004)]. This probability is calculated each year, applying the same, constant coefficients [estimated by Beck (2004) on a different set of data]. The model comprises age, years of CSS membership, number of supplementary insurance products, and premium relative to the market average as explanatory variables. Age has a negative effect, accounting for the decreased mobility of older individuals. Duration of membership also has a negative effect because loyal members tend to remain loyal. Insured with several supplementary insurance products have more difficulty switching because the open enrollment requirement holds only for the compulsory component of coverage. Legally, it is possible to buy compulsory and supplementary coverage from different insurers, but consumers are afraid of insurers haggling over their obligation to pay, which easily results in delayed reimbursement.[5]

2.5 Results

2.5.1 Risk Adjustment and the Distribution of Expected Individual Profits

The choice of the risk adjustment scheme has a strong impact on the distribution of predicted individual profits and losses. As profits are zero in a competitive market, $(1/n)\sum_{i=1}^{n} E[\pi_{i,j}]$ is zero in all scenarios. Without risk adjustment (j = 0), the distribution is heavily skewed to the right, and exhibiting a marked tail of very sizeable losses (Figure 2.1, tail cut at -20,000). With a risk adjustment scheme including

[5]Finally, a high CSS premium relative to the market average encourages consumers to switch. However, construction of this variable in the present context would have required modeling the premium development of competitors (which in turn would depend also on payments into and from the risk adjustment scheme). Therefore, this ratio is set equal to one to avoid this complication.

prior hospitalization and PCGs (j = 3) the distribution of $\pi_{i,3}$ is almost symmetrically centered at zero (Figure 2.2), with its median value equal to CHF 489, down from CHF 5,985 with no risk adjustment. However, there is a heavy tail of very profitable consumers, who (in expectation) seem to be overpaid by the risk adjustment scheme.

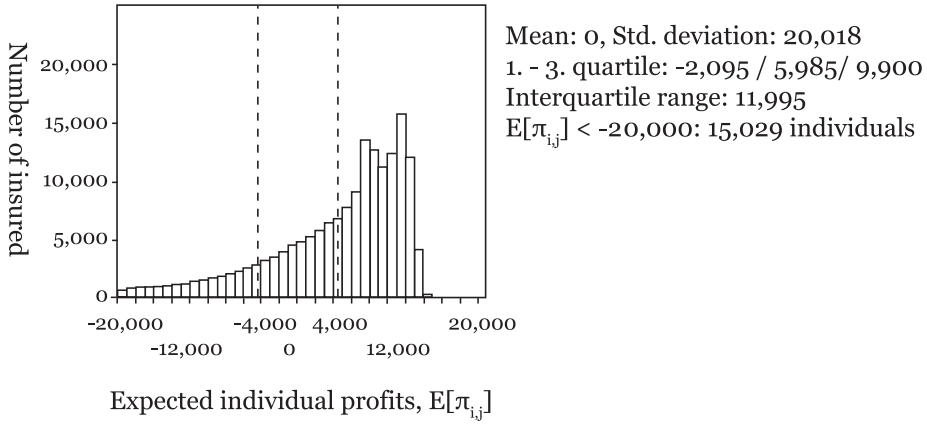

Figure 2.1: Expected Individual Profits Without Risk Adjustment (0).

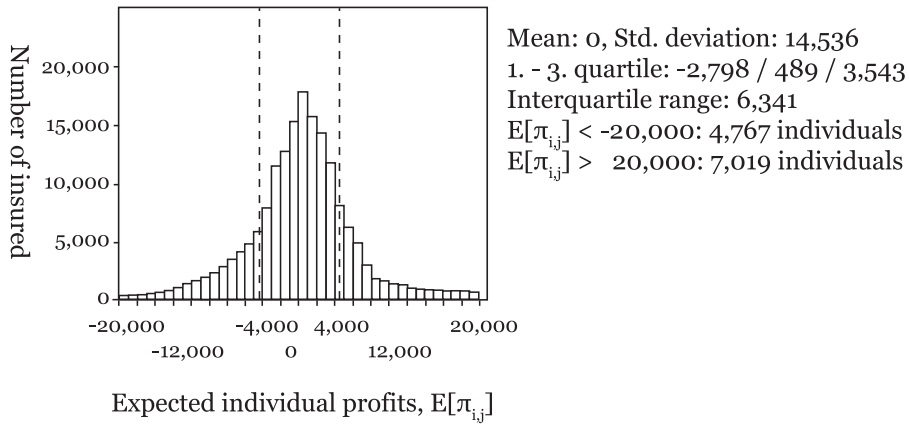

Figure 2.2: Expected Individual Profits With Risk Adjustment (3).

The objective of risk adjustment is to neutralize insurers' incentive for risk selection given that premiums are regulated not to reflect true risk. It therefore should affect the composition of the risk pool in terms of the subsets A through D distinguished in Section 2.2 above. Indeed, variant (3) causes the share of risks in subgroup A (expected individual profit $>$ CHF 1,000 p.a.) to drop from 56 percent (no risk adjustment) to a minimum of 20 percent. Since the left tail is not thinned out com-

pletely (compare Figures 2.1 and 2.2), the share of individuals in the unfavorable group D decreases only slightly from 21 to 18 percent. Table 2.5 exhibits the full set of estimates. However, as will be shown in the next section, risk adjustment causes the composition of these groups to change completely.

Table 2.5: Effect of Risk Adjustment on the Size of the Four Subgroups A – D.[a]

	Calculated population shares, in percent			
Risk adjustment formula No.	A	B	C	D
(0) None	56	14	9	21
(1) Age, gender	40	27	14	18
(2) Age, gender, prior hospitalization	26	34	23	17
(3) Age, gender, prior hospitalization, PCG	20	35	27	18

[a] A: $\pi_{i,j}$ >1,000; B: 0<$\pi_{i,j}$<1,000; C: 0>$\pi_{i,j}$> -1,000; D:< -1,000

2.5.2 Effect of Risk Adjustment on the Characteristics of Subgroups

Risk adjustment may have an important effect on the characteristics of the subgroups making up the risk pool. This effect seems to have been largely neglected in the literature. However, the political acceptance of a risk adjustment scheme strongly depends on its distributional impacts. Since incomes are not known, the analysis of this section is limited to age and two indicators of health status, viz. HCE prior to risk adjustment and membership in one or more PCG.

Without risk adjustment (0), profitable members (A) are pretty much the usual suspects, viz. younger than average, low net HCE and rare PCG membership, indicating absence of chronic illness (see Table 2.6). Unsurprisingly, demographic risk adjustment (1) changes this picture strongly in terms of age, causing mean age in subgroup A to increase from 46 to 62 years. Conversely, average age drops by as much as 20 years in subgroups B and C while it still decreases by 11 years in subgroup D. At the same time, HCE (net of cost sharing) more than doubles in subgroup A, while membership in at least one PCG more than triples, reaching 11 percent (not far below the overall share of 15 percent). This indicates that even mere demographic risk adjustment can turn some chronically ill individuals into profitable customers.

However, the D subgroup continues to have by far the highest average HCE and share of PCG members.

Table 2.6: Effect of Risk Adjustment (RA) on Characteristics of Subgroups.[a]

Risk Adjustment formula No.	Mean age in category, 2000 Overall mean: 54				HCE per month, 2000 (CHF, prior to RA) Overall mean: 291				Share of individuals in $\geq$1 PCG, 2000 (percent) Overall mean: 15			
	A	B	C	D	A	B	C	D	A	B	C	D
(0) None	46	59	64	67	77	211	320	910	3	16	27	40
(1) Age, gender	62	44	44	56	175	130	197	856	11	8	14	33
(2) Age, gender, prior hospitalization	70	49	45	51	328	138	176	713	20	8	11	26
(3) Age, gender, prior hospitalization, PCG	71	51	46	51	441	132	159	631	47	7	4	10

[a] A: $\pi_{i,j}$ >1,000; B: 0<$\pi_{i,j}$<1,000; C: 0>$\pi_{i,j}$> - 1,000; D:< -1,000

When prior hospitalization is included in the Type (2) risk adjustment scheme, characteristics of the subgroups change again. Average age in the A group even increases to 70, exceeding the value implied by the demographic formula (1). Because morbidity is higher among the elderly, they are more likely to get a morbidity related subsidy, making them attractive risks to the insurer. The change in the composition with regard to health status is even more striking. Average net HCE in the subgroup A now is CHF 328, exceeding the overall mean of CHF 291 per month. Likewise, A now contains 20 percent individuals who are in some PCG (overall mean, 15 percent). Put the other way round, many ill people may now contribute to expected profit even if their expected HCE is above average, while the very healthy are transformed into average risks because they are loaded with payments to the risk adjustment scheme.

When PCGs are also included into the scheme (Type 3), average age remains roughly the same in all subgroups. However, average net HCE in the A group increases, while those of the unfavorable D group decrease once more, to CHF 631 per month (which is still more than twice the overall mean of CHF 291). The most amazing change, however, is in PCG membership. The A subgroup now consists of 47 percent chronically ill.

In sum, this analysis offers important insights into the workings of risk adjustment Schemes (2) and (3). By collecting transfers from healthy people, they transform them (on average) from very favorable into medium risks from the insurer's point of view. The high risks on the other hand come with a subsidy, making some of them very profitable. These profits indicate an overpayment by the risk adjustment scheme, which can be explained as follows. While the subsidy equalizes average expected contributions across risk groups, a majority of individuals has HCE below group average because the distribution of HCE is skewed to the right even within a risk group, with a long tail towards high values.

Before turning to the financial benefits attainable by risk selection, two points should be mentioned. First, a morbidity-based risk adjustment makes it more difficult to recognize risks. With no or only demographic risk adjustment, it is sufficient for insurers to gather information about prior utilization and personal well being, whereas with Schemes of Type (3), they will have to establish precisely those chronic conditions that yield the highest contributions - a far more complex task. The second point concerns risk selection through quality of services covered. As pointed out by e.g. Newhouse (1982) and Van de Ven and Ellis (2000), insurers may try to stave off unfavorable risks by e.g. incorporating lower quality care for the chronically ill (always on the premise of community rating combined with imperfect risk adjustment). This form of selection hurts some of the weakest. Moreover, it is difficult to contain because service quality cannot easily be regulated [Marchand et al. (2004)]. The authors cited agree in their expectation that morbidity-related risk adjustment of the Types (2) and (3) is a suitable tool to prevent this type of selection. The present analysis supports this notion by showing that Scheme (3) causes the most favorable risk group to contain a large number of chronically ill individuals, who are eschewed by insurers who offer them benefits of lower quality. Conversely, insurers with a small market share (and therefore little influence on mean HCE) may benefit by attracting the chronically ill through special programs. In the Netherlands, where morbidity

indicators are included in the risk adjustment scheme, an insurer in fact developed a disease management program for diabetes patients [Van de Ven et al. (2007)].

2.5.3 Estimating the Financial Incentive for Risk Selection

In this section, we derive an estimate of the financial gains from risk selection by calculating ex-post profits associated with the strategies described. According to Equation (2.2), these profits are a function of the risk adjustment Schemes (0) to (3). The financial advantage due to risk selection is expressed as the ability to offer a lower premium, which almost certainly leads to a favorable market position in view of strong price competition. From the moment an insured in our sample dies or switches to another insurer, her cash flow drops to zero, as in real life. All figures are in prices of 2000.

Table 2.7: Premium Reduction Thanks to Deterring D-rated Risks, in Percent of Average Premium.

Risk adjustment formula No.	Possible premium reduction
(0) None	46
(1) Age, gender	32
(2) Age, gender, prior hospitalization	19
(3) Age, gender, prior hospitalization, PCG	16

First, the strategy of deterring expectedly unfavorable risks (D) is evaluated. An insurer capable of getting rid of all expectedly unfavorable risks (D) could reduce its average premium level by as much as 46 percent on expectation when there is no risk adjustment at all (see Table 2.7). Introducing age and gender as risk adjusters already serves to reduce the potential for premium reductions by one third. The most elaborate variant (3) achieves a reduction of two thirds, to 16 percent.

The other strategy is to attract risks that are expected to be favorable (A). Let x denote the factor by which the number of A-rated customers is increased, the number of risks in all other subgroups held constant. If x goes to infinity, the risk portfolio consists to over 99 percent of A-rated customers. Because many Swiss insurers are

rather small, $4 \leq x \leq 6$ are considered realistic values. Without risk adjustment, this strategy on expectation permits a lowering of the premium level by 48 percent (x=6, in Table 2.8). Demographic adjustment (1) reduces this figure to 38 percent. However, it takes Schemes (2) and (3) involving prior hospitalization and PCGs to largely reduce the financial gains associated with 'chasing the good risks'.

The entries of Tables 2.7 and 2.8 indicate that even in the long-run, 'regression to the mean' does not equalize risk profiles. Without risk adjustment, both deterring unfavorable and attracting favorable risks are highly profitable strategies. However, adding the crude morbidity indicator 'prior hospitalization' works surprisingly well to neutralize these gains. A possible explanation is the fact that hospitalization is a good proxy for HCE in earlier years, which are used for prediction by the insurers.[6] It is well known that risk adjustment schemes work best if they predict cost to the same degree as insurers themselves.

Table 2.8: Premium Reduction Thanks to Attracting A-rated Risks, in Percent of Average Premium.[a]

Risk adjustment	x = 2	x = 4	x = 6	x = 8
(0) None	23	41	48	66
(1) Age, gender	17	32	38	57
(2) Age, gender, prior hospitalization	7	15	19	31
(3) Age, gender, prior hospitalization, PCG	6	14	18	34

[a] x: Intensity of risk selection (x=2→number of A-rated risks is doubled)

The contribution of PCGs on the other hand is somewhat disappointing, raising doubts whether the administrative expense for establishing them is worthwhile. However, Table 2.6 shows that the risk adjustment formula (3) including PCGs excels in directing subsidies specifically to individuals with consistently high HCE. The incentive to skimp on the quality of care for the chronically ill is certainly the weakest with this formula.

As shown in Table 2.9, misclassification is likely to occur, its extent clearly depending on the risk adjustment scheme implemented. For example, the fraction of

[6] We thank Erik Schokkaert for pointing this out.

individuals rated A who ex-post generated losses is 7 percent without risk adjustment but 25 percent with risk adjustment (3). In other words, with risk adjustment (3) in place, as many as one-fourth of all seemingly very attractive customers were misclassified. For the D-rated customers, the frequency of misclassification is quite high for all risk adjustment schemes. This likely reflects the fact that – due to the lack of diagnostic information – predictions were based on total prior cost, failing to distinguish between acute and chronic illnesses. Again, false classifications occurred more frequently with a morbidity-based risk adjustment scheme in place. These results confirm our previous claim (Section 2.5.2) that a morbidity-based risk adjustment scheme makes it more difficult to distinguish between profitable and unprofitable risks. In the absence of risk adjustment, unprofitable risks are simply all customers with above-average HCE, for example all chronically ill individuals. By subsidizing some conditions leading to high HCE, morbidity-based risk adjustment changes this, turning some of the high risks into very profitable customers (see Table 2.6). In sum, morbidity-based risk adjustment requires a much more careful screening of individual risk profiles and health conditions by insurers seeking to pick profitable risks.

Table 2.9: Frequency of Misclassification.

Risk adjustment formula No.	Percentage of A-rated customers who ex post turn out to generate losses	Percentage of D-rated customers who ex post turn out to generate gains
(0) None	7	24
(1) Age, gender	15	28
(2) Age, gender, prior hospitalization	23	35
(3) Age, gender, prior hospitalization, PCG	25	38

2.5.4 Variation of the Planning Horizon

As stated in the Introduction, it is reasonable to assume that insurers plan over a period of several years. In order to assess the importance of this consideration, Equations

(2.1) to (2.4) were recalculated assuming planning horizons of 4, 3, 2, and 1 year.[7] The calculation of these models is straightforward. In expressions (1) – (4), the summation stops at the pertinent year prior to 2004. To estimate expected HCE, Formula (2.5) is applied replacing $E[HCE_{i,2004}]$ by $E[HCE_{i,t}]$, with $t \in 2000, 2001, ..., 2003$.

Table 2.10: Size of Subgroups A and D with Different Planning Horizons.

Risk adjustment formula No.	Percentage A-rated			Percentage D-rated		
	1 year	3 years	5 years	1 year	3 years	5 years
(0) None	70	68	56	18	18	21
(1) Age, gender	52	49	40	19	19	18
(2) Age, gender, prior hospitalization	44	39	26	18	18	17
(3) Age, gender, prior hospitalization, PCG	40	34	20	18	18	18

With a shorter planning horizon, a considerably higher share of individuals were classified as very profitable A-types (see Table 2.10). For example, with a 1-year planning horizon and risk adjustment (3), 40 percent of customers would have been classified as A as opposed to 20 percent with a 5-year planning horizon. This is likely due to the 'regression to the mean' specification of the forecasting model. For instance, let an elderly lady have small observed HCE in 1999. Then, her forecast HCE for 2000 would be quite low as well according to Equation (2.5). Over the years, however, predicted values approach the high (conditional) mean HCE of a person of her type. The other components of Equation (2.1) being largely independent of the planning horizon, her rating might well drop from A in the 1-year analysis to a B or C in the longer term.

At the other end, however, the number of D-rated customers does not vary much with the planning horizon (see Table 2.10). In fact, those who had high HCE in 1999 are mostly predicted to have high HCE in all subsequent years. This is especially true for those with high drug or long-term care expenditure, which point to chronic illness (see Table 2.4). Indeed, the entries of Table 2.4 suggest the existence of a basically

[7] We thank Mathias Kifmann, our discussant at the 19th annual meeting of the Health Economics Working Party within the Verein für Socialpolitik, and an anonymous referee for pointing this out.

Table 2.11: Frequency of Misclassification, Different Planning Horizons.

	Percentage of A-rated customers who ex post turn out to generate losses			Percentage of D-rated customers who ex post turn out to generate gains		
Risk adjustment formula No.	1 year	3 years	5 years	1 year	3 years	5 years
(0) None	10	11	7	33	31	24
(1) Age, gender	10	15	15	33	31	28
(2) Age, gender, prior hospitalization	13	21	23	43	40	35
(3) Age, gender, prior hospitalization, PCG	15	21	25	45	41	38

healthy type (with a low autoregressive element in HCE) and a chronic type (with a high element). This difference has implications for the risk of misclassification as a function of the planning horizon (see Table 2.11).

Table 2.12: Premium Reductions Thanks to Risk Selection, Different Planning Horizons (Percent of Average Premium).

Risk adjustment formula No.	Gains from attracting A, x= 4			Gains from deterring D		
	1 year	3 years	5 years	1 year	3 years	5 years
(0) None	37	35	41	45	40	46
(1) Age, gender	35	31	32	38	33	32
(2) Age, gender, prior hospitalization	18	17	15	19	19	19
(3) Age, gender, prior hospitalization, PCG	18	17	14	17	17	16

Considering A-rated customers first, their low autoregressive element in HCE causes predictability to decrease with length of planning horizon in principle [see rows (1) through (3)]. The countervailing effect is regression to the conditional mean, which may make for more accuracy of prediction if the insurer selects risk based on observable characteristics unfettered by risk adjustment [see row (0)]. Turning to D-rated individuals, their high HCE is the consequence of a whole host of causes, resulting in a low degree of predictability and hence high frequency of misclassification. However, this group comprises an important share of chronically ill with their marked element of autoregression in HCE. This may explain why the risk of misclas-

sification, while high compared to A-rated customers, decreases with length of the time horizon.

Finally, Table 2.12 displays the profits associated with risk selection as function of the planning horizon. Starting with the strategy of attracting A-rated risks, the gains tend to increase, which accords with the finding that the frequency of misclassification decreases in the absence of risk adjustment [row (0) of Tables 2.11 and 2.12]. By way of contrast, refinements of the risk adjustment formula cause gains not only to be lower but to decrease with the planning horizon [rows (1)–(3), again in keeping with Table 2.12]. In the case of D-rated customers, gains have a similar pattern [row (0) of Table 2.12], likely reflecting the tendency for predictability of HCE to improve [row (0) of Table 2.11]. Again, these gains are reduced strongly by refinements of the risk adjustment formula.

Reading Table 2.12 vertically column by column, one sees that the gains from attracting A-rated or deterring D-rated customers can be reduced by at least one-half by the more refined risk adjustment formulas (2) and (3). Therefore, the relative effectiveness of the risk adjustment formulas neither hinges on the insurer's choice of planning horizon nor on the length of its planning horizon.

2.6 Conclusions

There is a broad consensus that given managed competition with community-rated premiums, risk adjustment becomes a necessary regulation of health insurance markets. However, while a purely demographic risk adjustment formula has been recognized as being insufficient, its precise specification has remained controversial. Most of the empirical literature describes and tests for the relationship between a set of morbidity indicators and HCE of 1 year. This short time horizon is in accordance with the fact that managed competition usually allows for annual open enrollment. However, insurers have a strong interest in long-term customers in view of considerable acquisition cost. Moreover, consumers likely would think twice before signing

up with an insurer whose planning horizon is as short as 1 year. Given a longer term perspective, insurers' strategies are influenced by two empirical facts, viz. decreasing precision of forecasts and regression to the (conditional) mean. The first fact causes the risk of misclassification to increase with a longer planning horizon. However, to the extent that the regression is to the conditional rather than the grand mean of HCE, the second fact may serve to increase the payoffs to risk selection when the planning horizon is extended.

In this research, the effectiveness of risk adjustment schemes is assessed in the light of these considerations. Purely demographic risk adjustment already increases the likelihood of misclassification but does not sufficiently neutralize the longer term, systematic differences in HCE to decisively mitigate the incentives for risk selection. For this, it takes a more refined formula that includes prior hospitalization and possibly Pharmaceutical Cost Groups (PCGs) as risk adjusters.

Gains from risk selection are estimated based on predicted HCE net of copayments, premiums, and risk adjustment payments, discounted to present value and weighted by the probabilities of death and of switching to a competitor to obtain expected values. These calculations are performed for four different risk adjustment schemes, viz. (0) no risk adjustment, (1) demographic risk adjustment, (2) demographic risk adjustment, with prior hospitalization added as a simple morbidity indicator, and (3) PCGs complementing Scheme (2).

For a sample of some 180,000 Swiss individuals, expected net present values conditioned on the risk adjustment scheme were calculated. However, these values must exceed the variable cost associated with risk selection effort in order to trigger action on the part of the insurer. Since this cost is unknown, an arbitrary but not unrealistic value of $\pm$CHF 1000 p.a. ($ 800 as of 2007) serves as a threshold. Thus, the insurer is assumed to be indifferent with regard to risks whose contributions to expected profit fall within this interval.

The risk adjustment schemes distinguished modify incentives for risk selection according to expectations. The better they reflect morbidity, the smaller the share of

the insured population that constitutes favorable and unfavorable risks, respectively. Adjustment using only age and sex (Type 1) causes the share of favorable risks to drop from 56 to 40 percent, the share of unfavorable ones, from 21 to 18 percent. With schemes of Type (2) and (3), the figures for the favorable risks drop to 26 and 20, and for the unfavorable risks, to 17 and 18 percent, respectively. It also should be noted that the characteristics of the subgroups change dramatically with type of risk adjustment. Average age of favorable customers increases from 46 (Type 0) to 71 years (Type 3), while the share of those belonging to one or more PCGs (an indicator of chronic illness) increases from 3 to 47 percent.

The success of risk selection efforts is reflected by the insurer's ability to lower premiums and hence gain market share. In the absence of risk adjustment, deterring all unprofitable risks is estimated to result in a 46 percent premium reduction over 5 years. This longer term competitive advantage is reduced to 16 percent under Type (3) risk adjustment, which takes into account both prior hospitalization and membership in a PCG. Interestingly, this figure is in the same range as the premium reductions that may be offered for participation in a managed care alternative, which constitutes a product innovation.

Thus, it may be argued that Type (3) risk adjustment is effective enough to redirect insurers' efforts from risk selection to product innovation. In addition, the risk of misclassification is a mere 7 percent for an insurer 'chasing the good risks' as long as there is no risk adjustment but increases to 25 percent under Scheme (3). Refined risk adjustment therefore becomes even more effective than indicated by the expected contribution to profit because it exposes insurers to increased uncertainty.

In conclusion, this research suggests that risk adjustment can be designed in a way as to be effective enough in the longer term to redirect insurers' efforts away from risk selection in favor of product innovation while using easily available information such as prior hospitalization and membership in pharmaceutical cost groups.

Acknowledgments

An earlier version of this paper was presented at the 7th conference of the European Risk Adjustment Network (RAN), March 2006 in Berlin, Germany. The authors thank the members of the RAN, especially Wynand van de Ven and Erik Schokkaert, and two anonymous referees for their helpful comments.

Supply-side and Demand-side Cost Sharing in Deregulated Social Health Insurance: Which Is More Effective?

Maria Trottmann, Peter Zweifel, and Konstantin Beck

Abstract: Microeconomic theory predicts that if patients are fully insured and providers are paid fee-for-service, utilization of medical services exceeds the efficient level ('moral hazard effect'). In Switzerland, both demand-side and supply-side cost sharing have been introduced to mitigate this problem. Analyzing a panel dataset of about 160,000 adults, we find both types of cost sharing to be effective in curtailing the use of medical services. However, when moral hazard mitigation is traded off against risk selection, the minimum-deductible, supply-side cost sharing option ranks first, followed by the medium-deductible demand-side alternative, making the supply-side option somewhat more effective.

Keywords: Health insurance, moral hazard, copayment, two-stage residual inclusion
JEL classification: I11; G22; D82

Accepted for publication in 'Journal of Health Economics'.

Chapter 3

Supply-side and Demand-side Cost Sharing in Deregulated Social Health Insurance: Which Is More Effective?

3.1 Introduction

One of the main goals of health care financing systems is to promote efficient levels and types of care [Ellis and McGuire (1993)]. If patients are fully insured and providers are paid fee-for-service, they desire larger than optimal quantities of health care services, connoted 'moral hazard'. Zeckhauser (1970) and Zweifel and Manning (2000) have analyzed how demand-side cost sharing (in the guise of deductibles or co-payments) can be used as a corrective. However, demand-side cost sharing exposes consumers to financial risk, contradicting the very objective of insurance. Unless limited by a stop-loss, it also makes beneficial procedures unaffordable to some patients [Nyman (1999)]. In addition, it might be considered unfair towards the chronically ill.

These considerations have created interest in the alternative of supply-side cost sharing (in the guise of capitation or prospective payment). Because of their information advantage, providers of medical care can influence the demand for their services

to a greater extent than other professionals [Arrow (1963)]. Moreover, providers are less vulnerable to risk than patients because they can pool treatment cases. However, supply-side cost sharing might also promote a reduction in quality or the denial of beneficial but costly services, a phenomenon commonly termed stinting [Newhouse (2002)].

Both demand-side and supply-side cost sharing have been empirically examined in terms of their effectiveness. The novelty of this paper is that it directly compares the expenditure effects of demand-side and supply-side cost sharing (and combinations thereof), using contract variants offered by the same health insurer. This has the advantage that many side conditions (underwriting policy, billing procedure) are kept constant. Moreover, the paper complements Lehmann and Zweifel (2004), who construct a proxy for unobserved health status from prior health care expenditure (HCE), by the two-stage residual inclusion estimator [2SRI, Terza et al. (2008)]. In this way, self-selection effects are more fully controlled for. Finally, it extends the set of instruments influencing choice of plan but not HCE by including the premium for the baseline contract, the potential premium reduction for a restricted plan, the individual's credit record, and years of membership with the same fund. For the capitated plan, an additional instrument is a dummy indicating whether or not an independent practice association (IPA) was operative in the individual's county of residence.

The data comes from Switzerland, a country where consumers have annual free choice of plan with no employer involvement. The chronically ill are not precluded from switching due to open enrollment. Low-income individuals (about 30 percent of the population) are eligible for premium subsidies. Receiving the subsidy, they are less likely to chose high-deductible plans, because a reduction in the income transfer to the sick state is particularly disadvantageous if the income effect on medical consumption is strong [Nyman (1999)]. On the other hand, managed-care type plans that are too restrictive compared to premium charged will not be chosen.

The remainder of this article is structured as follows. Section 3.2 contains an overview of the empirical literature. The policy setting is described in Section 3.3, while Section 3.4 is devoted to a description of the data base. In Section 3.5, we explain the econometric methods used to separate moral hazard from self-selection effects and to deal with the very skewed distribution of the HCE data. The estimation results are presented in Section 3.6. Section 3.7 discusses policy implications in view of related literature, while the final Section 3.8 contains a summary and conclusions.

3.2 Literature Review

In order to keep this review concise, we focus on empirical papers that measure moral hazard in health insurance. When individuals have a choice of plan, self-selection effects need to be accounted for because those who expect high future HCE are more likely to opt for more comprehensive insurance. A small number of researchers have avoided this selection problem by benefiting from randomized experiments [the famous RAND study; Manning et al. (1987)] or natural experiments [Chiappori et al. (1998), Eichner (1998), or Winkelmann (2004)]. Other papers have used econometric techniques to address endogenous plan choice. Many econometric approaches require for identification the availability of at least one variable that influences contract choice but not utilization (an 'identifying instrument'). Pertinent studies from Switzerland are Schellhorn (2001), Gerfin and Schellhorn (2006) and Gardiol et al. (2006). The former two rely on premium level and supplementary hospital insurance as identifying instruments, while the latter uses death as an indicator of morbidity which is unaffected by insurance. Using Australian data, Cameron et al. (1988) advocate income as determinant of insurance coverage but not utilization. In the United States, employers play a strong role in determining the individual's choice of plan, making their characteristics potential identifying instruments. For example, Dowd et al. (1991) and Cardon and Hendel (2001) exploit the fact that different employers offer different premiums and copayment levels, while Deb and Trivedi (2009) use the

employer's type (public or private), the size of the firm, and whether or not it offers both HMO and non-HMO options.

Turning to estimation techniques, one notices that instrumental variable estimators are rarely applied to non-linear frameworks with health data. An early exception is Dowd et al. (1991), who estimate a Tobit model with a correction for selectivity [Lee (1978)]. In addition, Deb and Trivedi (2009) and Deb et al. (2006) specify a fully parametric model of both choice and utilization equations, which is jointly estimated by maximum simulated likelihood. However, these approaches depend upon restrictive distributional assumptions. As HCE data is very skewed and the distribution of the 'tail' is difficult to specify correctly, Terza et al. (2008) advocate the two-stage residual inclusion estimator. It yields consistent estimates in a wide range of non-linear specifications.

Studies that have addressed endogeneity in non-linear panel data models are even more rare. Exemptions include Vella and Verbeek (1998) and Vella and Verbeek (1999), who propose a decomposition of the reduced form residuals into a part that drives the endogeneity and a random part. However, their approach requires strict distributional assumptions and is computationally intensive so it might not be applicable to very large data sets.

An alternate approach of exploiting the information of panel data was pioneered by Wolfe and Godderies (1991). It uses HCE from prior years to proxy unobserved differences between individuals which become predetermined in the year when the comparison between plans is performed [Lehmann and Zweifel (2004), Van Kleef et al. (2008)]. In this paper, a combination of the IV and the 'health proxy' approach will be applied.

3.3 Swiss Health Insurance

Swiss health insurance is of the 'managed competition' type (see Kreier and Zweifel (2010) for a comprehensive description). Coverage is mandatory for a rather compre-

hensive 'basic' basket of medical services and pharmaceuticals, written by some 80 private, not-for-profit insurers competing in a regulated market. Free consumer choice of plan is a distinctive feature of the system. There is no pre-selection of plans by employers or government agencies. Insurers are obliged to accept all applicants during annual open enrollment periods. Premium subsidies for low-income individuals are funded out of general tax. Premiums can be differentiated by area of residence but not by health risk. Reductions are possible for young adults (19-25) and individuals who receive accident coverage through the employer.

In the baseline contract, insured individuals enjoy unlimited access to all licensed physicians and most hospitals in their region of residence. They face a minimum annual deductible of CHF 300 (some EUR 200 as of 2006) and a copayment rate of 10 percent up to a cap of CHF 700 (EUR 470) per year. Physicians in independent practice are reimbursed fee-for-service (FFS) according to an administered fee schedule that is collectively bargained between the providers' and the insurers' associations. Hospitals receive per diems for patients treated (the nation-wide introduction of a DRG system is scheduled for 2012). The cantons[1] finance hospital investment and one-half of operational cost. While this system is generally found to ensure access to comprehensive health care to all citizens, it is criticized for high and rapidly increasing HCE, lack of co-ordination between providers, and lack of information about quality and efficiency [OECD (2006)].

In response to these problems, insurers have been granted the right to offer managed-care type options (since 1994) and higher deductibles (since 1996) in return for lower premiums. However, policy makers feared that these options would attract low risks. In addition to a risk adjustment scheme based on age and sex, they imposed limits on possible premium reductions. For voluntary deductibles, these are fixed percentages of the base premium or 80 percent of the additional financial risk taken by the consumer (deductible minus 300), whichever is less. The eligible deductible levels are also regulated, as shown in Table 3.1. In managed-care type contracts, the insurer

[1] Switzerland is divided into 26 cantons, with population ranging from 1,307,600 (Zurich) to 15,500 (Appenzell i.R.), Source: Swiss Federal Statistical Office, www.bfs.admin.ch.

must prove that the reduction is justified by efficiency gains rather than self-selection effects. Furthermore, it must not exceed 20 percent during the first five years since the launch of the contract. The same deductible levels apply to managed-care type and FFS plans.

Table 3.1: Regulation of Deductibles and Maximum Premium Reductions, 2006.

Deductible level in CHF / year	300	500	1,000	1,500	2,000	2,500
Max reduction in percent of the base premium	-	5	15	30	38	43
Max absolute reduction: 0.8 * (Deductible − 300)	-	160	560	960	1,360	1,760

CHF 1 ≈ EUR 0.66.

3.4 Data

The data base consists of individual records of more than 160,000 Swiss adults insured by CSS, a major Swiss insurer, and covering the years 2003 - 2006. It includes age, gender, residential location, contract choice, and HCE. Individuals who were not observed over the entire four years are excluded from the analysis, with death constituting the main cause. While the deathbound are known to cause a considerable amount of HCE, they exhibit an idiosyncratic pattern of health care utilization [see Werblow et al. (2007)], justifying separate analysis. The influence of closeness to death also calls for exclusion of individuals who died during 2007, resulting in a panel comprising some 160,000 individuals.

The supply-side cost sharing variant analyzed here is an IPA. Similar to the United States, participating physicians (mainly general practitioners) are paid a risk-adjusted capitation payment designed to cover all services rendered or prescribed up to a threshold of CHF 10,000 per patient and year. Beyond that limit, the insurer reimburses 90 percent of cost, as calculated according to the FFS alternative. Capitation payments are adjusted for age, gender, deductible level, hospitalization during previous year, nursing home stay during previous year, and 21 pharmaceutical cost groups. The pharmaceutical cost groups are similar to those used in the Dutch risk adjustment scheme [Lamers and Van Vliet (2003)]. While the insurer does not impose

guidelines or utilization reviews, many networks run them internally, combined with quality monitoring by independent auditors in some cases.

Table 3.2 shows descriptive statistics according to contract choice as of 2006. For simplicity, deductibles are grouped into three categories (minimum: 300, medium: 500, high: $\geq$ 1,000 CHF per annum). More than 70 percent of those who chose a deductible in excess of CHF 500 opted for the CHF 1,500 level. The other high deductible levels were only chosen by a small number of individuals each, so including them separately would have resulted in unstable estimates. Furthermore, observed HCE across the high deductible levels appeared to be similar.[2]

According to Panel A of Table 3.2, buyers of the high-deductible FFS plans are younger and more likely to be male than those with a minimum deductible. Their mean HCE amounts to CHF 1,057 or 23 percent of the CHF 4,610 pertaining to individuals with the minimum deductible. Also, their fraction of reporting positive HCE is 57 rather than 88 percent. If only those with positive HCE are taken into account, the mean is CHF 1,804 or 34 percent of the minimum-deductible benchmark of CHF 5,230, respectively. These differences point to sizeable effects of demand-side cost sharing (which still need to be corrected for self-selection effects, see below).

Turning to the supply-side cost sharing alternative (Panel B of Table 3.2), one notices that the HCE values for the IPA plans are lower throughout than for the conventional FFS plans with the same deductible level. In the minimum deductible group, average age is similar in the IPA and in the FFS plan. However, the high-deductible IPA variant is again characterized by a comparatively low mean age and

[2]For patients with high deductibles, it is questionable how well their HCE are observed. In earlier work [e.g. Lehmann and Zweifel (2004)], only individuals with the minimum deductible were analyzed on the grounds that patients have no incentive to submit their claims unless HCE exceeds the deductible. However, with the advent of electronic billing systems, the lion's share of billings are now transmitted directly from providers to insurers, who then charge the deductible to the patient. In some cantons, physicians even decided to abandon direct-to-consumer billings completely. CSS conducted an internal study relating the share of direct-to-consumer billing to HCE below the deductible. Contrary to expectations, the billing mode interacted with the deductible level had no influence on the probability of reporting positive HCE. It also had no significant impact on explaining positive HCE. It appears that many individuals submit their bills regardless of their deductible, maybe to decrease the administrative burden in case of an illness. In the dataset, 24 percent of individuals with positive HCE had HCE below their deductible.

Table 3.2: Descriptive Statistics According to Type of Contract, 2006.

Contract	N	Age	Male	HCE	Share with HCE > 0	HCE if >0
A. FFS plans						
Minimum DED (baseline)	84,053	55	0.4	4,610	0.88	5,230
		(18)	(0.49)	(8,961)	(0.33)	(9,378)
Medium DED	31,573	54	0.43	3,229	0.81	3,908
		(16)	(0.49)	(6,962)	(0.38)	(7,522)
High DED	38,386	45	0.54	1,057	0.57	1,804
		(13)	(0.5)	(3,593)	(0.49)	(4,580)
B. IPA plans						
Minimum DED	4,942	54	0.44	2,933	0.85	3,427
		(17)	(0.5)	(5,999)	(0.35)	(6,355)
Medium DED	1,134	49	0.45	1,686	0.78	2,121
		(15)	(0.45)	(3,654)	(0.41)	(3,999)
High DED	3,598	43	0.56	834	0.58	1,415
		(12)	(0.56)	(3,431)	(0.49)	(4,368)

Standard deviations in parentheses, N = 163,686, DED = Deductible, IPA = Independent Practice Association, CHF 1$\approx$ EUR 0.66.

a higher share of men. Average HCE is CHF 834 or 18 percent of the minimum-deductible, FFS benchmark of CHF 4,610. The share of individuals with positive HCE is 58 percent rather than 88 percent, while mean HCE conditional on being positive amounts to CHF 1,415 or 27 percent of the CHF 5,230 benchmark.

In order to get a preliminary indication of the extent to which the cost differences may be caused by self-selection effects, it is instructive to compare the HCE of consumers who switch to higher deductibles and IPA plans with the HCE of those who do not (see Table 3.3). The switchers from a minimum-deductible FFS plan in 2005 to a high-deductible one in 2006 had caused HCE of CHF 826, a mere 19 percent of the non-switchers. Those changing to a medium-deductible alternative had caused HCE amounting to CHF 1,916 in 2005, or 44 percent of the non-switchers. Switchers who moved from a FFS to an IPA plan had HCE amounting to CHF 1,597, or 49 percent of the stayers. These figures point to substantial self-selection effects in both demand-side and supply-side cost sharing.

Table 3.3: Prior-year Mean HCE of Switchers and Non-switchers

Switch at the start 2006 HCE in 2005	FFS, Minimum DED to Medium DED	FFS, Minimum DED to High DED	FFS (all DED) to Capitated IPA
Non-Switchers	4,315	4,315	3,230
Switchers	1,916	826	1,597

DED = Deductible, IPA = Independent Practice Association, FFS = Fee-For-Service. CHF 1≈ EUR 0.66.

3.5 Econometric Model

3.5.1 Developing a Proxy for Unobserved Health Status, 2003 - 2005

The dataset does not contain direct information on health status such as diagnostic codes, restrictions on activities of daily living, or self-reported health. However, panel data allows to develop an indicator of health status from prior HCE [see Van Kleef et al. (2008), Lehmann and Zweifel (2004), Wolfe and Godderies (1991)]. In particular, Lehmann and Zweifel show how residuals from a random-effects Tobit regression of prior HCE on exogenous variables can serve as a proxy for unobserved health.

However, in view of considerable heteroscedasticity in the dataset, the two-part model is preferable over the Tobit. The first part is a random-effects probit model predicting the probability of observing positive HCE for individual i in year t [see Equation (3.1)]. The second part estimates the amount of HCE given that it is positive. The log transformation serves to reduce the skewness of the dependent variable. The present panel is unbalanced, as many individuals had positive HCE in some but not in all years. A Wooldridge test of serial correlation in the error term [Wooldridge (2002)] rejected the null hypothesis of no autocorrelation. Therefore, the feasible generalized least squares procedure proposed for unbalanced panels by Baltagi and Wu (1999) is applied to gain efficiency while avoiding biased estimation of standard errors. The model for deriving the health status proxy thus reads (all error terms assumed normally distributed),

$$Pr(HCE_{i,t} > 0) = \Phi(a + \beta X_{i,t} + \alpha_i + \upsilon_{i,t}) \tag{3.1}$$

$$log(HCE_{i,t}|HCE_{i,t} > 0) = b + \theta X_{i,t} + \gamma_i + \epsilon_{i,t} \tag{3.2}$$

$$\text{with} \qquad \epsilon_{i,t} = \rho * \epsilon_{i,t-1} + \xi_{i,t}$$

Equations (3.1) and (3.2) are estimated on the first three years of the dataset, i.e. 2003 to 2005. Explanatory variables are age, age interacted with gender, urbanization, area of residence, and a year dummy to account for inflation. Estimation results are shown in Table 3.10 of the appendix. Deviations from the expected value of HCE are averaged over the three years in order to reduce the influence of transitory health shocks.[3]

3.5.2 Endogeneity of Contract Choice, 2006

Even if the proxy derived from Equations (3.1) and (3.2) controls for unobserved differences in current health status, there are additional unmeasured variables that may cause someone opting for the minimum deductible to have a great deal of HCE, resulting in an overestimation of moral hazard effects. Examples are private information about probabilities of future illness, general attitude towards medical care, and previous experience with the health care system. Ignoring these confounders will lead to omitted variable bias in the HCE equation. Terza et al. (2008) show that the residuals from an equation modeling contract choice are good estimators of these confounders. Therefore, these residuals are included in the HCE equation alongside observed contract choice and the proxy for latent health (two-stage residual inclusion estimation, 2SRI). The 2SRI method also yields consistent estimates if the HCE equation is nonlinear. However, it requires equations

[3]Note that while only individuals with the baseline contract are included here, estimated coefficients will be used to predict individual HCE for the whole sample. This has the advantage that the endogeneity of contract choice does not bias estimators. In order to test robustness, we also calculated the proxy including all individuals, with little effect on results.

for contract choice to be specified.

For identification, at least one explanatory variable in the contract choice equation must not appear in the HCE equation. Five such variables are available.

1. Baseline premium:[4] A high baseline premium increases the attractiveness of higher-deductible and IPA options. At the same time, there is little reason why premiums should influence health care consumption. The income effect of premiums is limited in the Swiss case because low-income individuals (some 30 percent of the population) are eligible for a premium subsidy. Moreover, preliminary estimations showed that premiums do not influence HCE when other factors are controlled for.[5]

2. Absolute premium reductions for a higher-deductible or an IPA option: While premium reductions make these contract options more attractive, they should not influence health care consumption for the same reasons as described in item no. 1 above. They were also found to be insignificant in an estimation of HCE.

3. Number of years of CSS membership:[6] Long-standing members are known not to switch contracts, making them less likely to opt for a higher-deductible or an IPA option. However, loyalty is negatively correlated with health status because consumers who develop chronic conditions face a premium hike if they sign up with another insurer for the supplementary component (which they usually prefer to have from the same insurer to avoid ambiguity as to responsibility for payment). Nevertheless, preliminary estimations showed it to be insignificant in the HCE equation when entered in combination with the health status proxy. It therefore qualifies as an identifying restriction.

[4] Although we use data from only one insurer, premiums differ between regions. In addition, young persons and individuals who have accident coverage through their employer are eligible for premium reductions.

[5] Premium levels were also used as identifying instruments in Schellhorn (2001).

[6] This variable is truncated at 1999 because retrieving data from earlier years is cumbersome. There was a change in IT architecture in 1998.

4. Dummy indicating a bad credit record: This may reflect lower income, which is relevant for contract choice because high deductibles are unattractive to risk-averse low-income individuals. At the same time, a bad credit record proved unrelated to HCE once the proxy for health status was included.

5. IPA officially on offer within the individual's area of residence: The availability of an IPA importantly favors the choice of the corresponding option. However, it proved to be unrelated to utilization provided regional differences were controlled for by dummies.

Modeling the choice of deductible calls for an ordered probit model, while for the choice of the IPA a probit model is sufficient. For the probit, the generalized residuals were derived by Gourieroux et al. (1987). Let h_i be an indicator variable equal to one if the IPA plan was chosen and zero otherwise, z_i a vector of covariates, and $\hat{\theta}$ a vector of the estimated coefficients. Then, the generalized residuals $\hat{u_i}$ are given by

$$\hat{u_i} = h_i * \frac{\phi(z_i'\hat{\theta})}{\Phi(z_i'\hat{\theta})} + [1 - h_i] * \frac{-\phi(z_i'\hat{\theta})}{1 - \Phi(z_i'\hat{\theta})} = \frac{[h_i - \Phi(z_i'\hat{\theta})]\phi(z_i'\hat{\theta})}{[1 - \Phi(z_i'\hat{\theta})]\Phi(z_i'\hat{\theta})}, \tag{3.3}$$

where Φ denotes the cumulative and ϕ, the standard normal density respectively. In the same spirit, the generalized residuals for multinomial or ordered choice models have been defined by Vella (1993). Let there be $i = 1...N$ individuals choosing from $k = 1...K$ ordered alternatives, and let d_{ik} denote an indicator function taking the value 1 if individual i has chosen alternative k and zero otherwise. Then, generalized residuals $\hat{v_i}$ are given by

$$\hat{v_i} = \Sigma_{k=1}^{K} d_{ik} \frac{\hat{\pi}_{ik}[d_{ik} - \hat{\Pi}_{ik}]}{[1 - \hat{\Pi}_{ik}]\hat{\Pi}_{ik}} \tag{3.4}$$

with $\hat{\Pi}_{ik}$ denoting the estimated cumulative probability that individual i chooses the k-th alternative and $\hat{\pi}_{ik}$, the estimated value of the density at that point. These two quantities are determined as follows. Let $\hat{\gamma}$ be the vector of estimated coefficients

from the ordered probit and $\hat{\alpha}_k$, the estimated cut points with $\hat{\alpha}_0 = -\infty$ and $\hat{\alpha}_K = \infty$. Then,

$$\hat{\pi}_{ik} = \phi(\hat{\alpha}_{k-1} - z_i'\hat{\gamma}) - \phi(\hat{\alpha}_k - z_i'\hat{\gamma}) \text{ and } \hat{\Pi}_{ik} = \Phi(\hat{\alpha}_{k-1} - z_i'\hat{\gamma}) - \Phi(\hat{\alpha}_k - z_i'\hat{\gamma}). \quad (3.5)$$

3.5.3 Specification of the Two-part Model, 2006

The distribution of HCE has a cumulation point at zero. Among the alternatives available for dealing with this fact, the two-part model is preferred over e.g. the Tobit model for two reasons. First, the zeroes are perceived as reflecting choices rather than missing values [see Jones (2000)]. Second, both supply- and demand-side cost sharing are known to affect the decision to use health care at all differently from the decision how much care to use.

The first part of the two-part model is often specified as a probit and estimated by maximum likelihood. However, with the inclusion of the residuals from the contract choice equations, its errors are non-normally distributed, causing maximum likelihood to be inconsistent.[7] To avoid this problem, a GLM estimation with a probit link is applied to the HCE data of 2006. This method is consistent as long as $E(l|x) = \Phi(x\beta)$, with l denoting an indicator that equals one if HCE > 0 and zero otherwise. Normal distribution of the error terms is not required [Cameron and Trivedi (2005)].[8]

The specification of the second part ($HCE|HCE > 0$) has been discussed by Manning (1998) and Manning and Mullahy (2001) (MM hereafter). Because of the positive skewness of the dependent variable, raw-scale estimates can be imprecise even in large datasets. The log transformation is often used to mitigate skewness, with coefficients interpreted as (semi-) elasticities of the mean response. However, Manning (1998) shows that if the error variance is heteroscedastic in a way that is

[7] The generalized residuals are non-linear transformations of normally distributed variables. These are not normally distributed. Then, the errors from the two-part model are linear combinations of several normally distributed variables and one non-normally distributed variable. This combination is not normally distributed.

[8] There is no contradiction to Terza et al. (2008), who suggested estimation by non-linear least squares. Although it goes by a different name, GLM is an iteratively reweighted nonlinear least squares estimator [Hardin and Hilbe (2007)].

correlated with the covariates, these coefficients are no longer consistent elasticity estimates. Moreover, a retransformation of predicted values is required as soon as absolute rather than relative savings due to cost sharing are of interest.

Blough et al. (1999), MM and others suggest estimating $ln(E[y|x, y > 0])$ directly by a GLM procedure with a log link (i.e. $ln(E[y|x, y > 0]) = x\beta$) and an appropriate variance function. As MM point out, the GLM estimates are consistent as long as the mean function is correctly specified, but might lead to imprecise estimates if the residuals are positively skewed even after transformation to log. Following the procedure for model selection suggested by MM, we start with a consistent GLM procedure, the gamma regression. The kurtosis of the 2006 residuals on the log scale is 3.53. This creates a tradeoff between imprecision (GLM) and possible bias (OLS applied to ln(y)). In this work, GLM is used because taking heteroscedasticity into account is deemed more important than precision. Given GLM, a Park test is performed to select the variance function. The estimated λ is 1.81, which is closest to the gamma specification.

The residuals from the contract choice equations are estimates rather than observations. Not accounting for this in the outcome estimation could lead to downward biased estimates of the standard errors [Heckman (1976)]. Therefore, the standard errors were obtained by bootstrapping.[9]

3.6 Results

3.6.1 Effects of Demand-side and Supply-side Cost Sharing

The results for the first part of the two-part model estimated on HCE data for 2006 are shown in the first three columns of Table 3.4. The first column pertains to the full model. The second column excludes the residuals from the contract choice equation. The third column corresponds to a naive specification that also excludes the proxy for

[9]The estimation was repeated 400 times after resampling with replacement (clustered by patient). In our specific application, the bootstrapped standard errors turned out to be similar to those obtained in the original estimation.

health status. For the variables of interest (DED, IPA), marginal effects are calculated for a representative individual (in italics below the coefficients), i.e. a woman at the age of 52, living in a suburban community in the Zurich region, having the baseline contract plus accident coverage, and a supplement covering alternative medicine. The health proxy is taken at its sample average. This individual's estimated probability of positive HCE is roughly 91 percent. For interaction terms, marginal effects are calculated according to the formulas provided by Norton et al. (2004).[10]

Voluntary deductibles are found to progressively reduce the probability of positive HCE; however, their incentive effect shrinks to 0.7 percentage points for the medium and 3.7 points for the high deductible category, respectively, according to the full specification. Quite generally, estimated values are about two to seven times smaller in the full and restricted than in the naive specification, pointing to considerable self-selection effects based on health status. On the other hand, they cannot be said to depend on the type of plan (DED * IPA insignificant except in the naive specification). By way of contrast, membership in an IPA may be associated with a higher probability of positive HCE (significant only in the restricted model), possibly reflecting preventive services offered. For example, one large IPA hands out vouchers for free immunizations against the flu in the fall.

Two groups of variables of Table 3.4 are worth a comment. First, supplementary hospital coverage (mainly for a private room) is associated with a higher probability of using medical services even after controlling for possible risk-selection effects. The same is true of supplementary coverage of therapies belonging to the category of alternative medicine. Second, the squared and the cubic form of the health proxy are highly significant as well. The impact of past, time-invariant health status on current expenditure thus does not appear to be linear over the whole distribution of HCE.

[10] To be specific, let β_a and β_b be the coefficients of two dummies, β_{ab} the coefficient of their interaction and $\bar{x}'\beta$ the influence of all other variables at representative values. The marginal effect of the interaction term is $\Phi(\beta_a+\beta_b+\beta_{ab}+\bar{x}\prime\beta)-\Phi(\beta_a+\bar{x}\prime\beta)-\Phi(\beta_b+\bar{x}\prime\beta)+\Phi(\bar{x}\prime\beta)$. For a dummy without interaction, the marginal effect is $\Phi(\beta_a+\bar{x}\prime\beta)-\Phi(\bar{x}\prime\beta)$. As these marginal effects are combinations of all coefficients, their standard errors are calculated by the delta method. The calculations are run in STATA using the nlcom command.

Table 3.4: Estimation Results from the Two-part Model, 2006.

	P(HCE>0), GLM with probit link			HCE|HCE>0, GLM with log link		
	Full	Restricted	Naive	Full	Restricted	Naive
Affluent community	0.039	0.041	0.056**	0.023	0.025	0.044
	(0.026)	(0.023)	(0.020)	(0.028)	(0.030)	(0.029)
Regional center	-0.052**	-0.052**	-0.036*	-0.092**	-0.090**	-0.084**
	(0.022)	(0.019)	(0.017)	(0.026)	(0.028)	(0.027)
Suppl. hospital	0.065***	0.069***	0.155***	0.030*	0.028	0.072***
	(0.011)	(0.011)	(0.010)	(0.014)	(0.015)	(0.014)
Suppl. altern. med.	0.036***	0.035***	0.085***	-0.076***	-0.078***	-0.063***
	(0.01)	(0.010)	(0.009)	(0.013)	(0.014)	(0.013)
Health proxy	0.473***	0.469***		0.523***	0.521***	
	(0.01)	(0.009)		(0.013)	(0.013)	
(Health proxy)^2	-0.061***	-0.061***		-0.085***	-0.085***	
	(0.004)	(0.004)		(0.007)	(0.006)	
(Health proxy)^3	0.005***	0.005***		0.003***	0.003***	
	(0.000)	(0.000)		(0.001)	(0.001)	
$\hat{v}_i$ DED	0.059			-0.332		
	(0.145)			(0.254)		
$\hat{u}_i$ IPA	-0.046***			-0.006		
	(0.013)			(0.019)		
Medium DED	-0.042**	-0.079***	-0.265***	-0.071***	-0.072***	-0.260***
	(0.016)	(0.012)	(0.010)	(0.014)	(0.016)	(0.015)
ME	*-0.007***	*-0.013****	*-0.055****			
	(0.003)	*(0.002)*	*(0.003)*			
Medium DED * IPA	0.084	0.083	0.122*	0.013	-0.001	0.053
	(0.066)	(0.061)	(0.054)	(0.071)	(0.078)	(0.075)
ME	*0.012*	*0.013*	*0.025***			
	*(0.008)***	*(0.008)*	*(0.011)*			
High DED	-0.207***	-0.295***	-0.869***	-0.182***	-0.191***	-0.837***
	(0.026)	(0.011)	(0.009)	(0.014)	(0.019)	(0.017)
ME	*-0.037****	*-0.054****	*-0.241****			
	(0.005)	*(0.003)*	*(0.005)*			
High DED * IPA	0.043	0.043	0.126***	0.066	0.054	0.188***
	(0.037)	(0.037)	(0.033)	(0.071)	(0.059)	(0.057)
ME	*0.011*	*0.014**	*0.034*****			
	(0.006)	*(0.006)*	*(0.009)***			
IPA	0.072	0.086**	-0.066**	-0.118*	-0.189***	-0.418***
	(0.042)	(0.028)	(0.025)	(0.059)	(0.035)	(0.033)
ME	*0.011*	*0.012****	*-0.012**			
	(0.006)	*(0.004)*	*(0.005)*			
AIC	0.676	0.676	0.885	18.088	18.088	18.365
N	163,686	163,686	163,686	128,744	128,744	128,744

*p<0.05, **p<0.01, ***p<0.001, DED = Deductible, IPA = Independent Practice Association, Standard errors in parentheses, ME = Estimated marginal effects. Additional regressors are age, gender, additional types of municipalities, region specific dummies, regulation on drug dispensing, accident coverage, long term care coverage, and youth rebate eligibility.

In the second part of the model, the amount of positive HCE is estimated (last three columns of Table 3.4). Higher voluntary deductibles are again found to progressively reduce HCE in the full and the restricted specifications, with 'true' savings due to incentive effects amounting to 7 percent (medium deductible) and about 18 percent (high deductible category), respectively. This time, the full and restricted models point to effects that are about four times smaller than according to the naive model. Type of contract does not play a role (DED * IPA insignificant except in the naive specification). Turning to supply-side cost sharing in the guise of an IPA, one notices a reduction of HCE amounting to 12 percent (full model) and 19 percent (restricted model) respectively, less than one-half of the 42 percent suggested by the naive specification.

3.6.2 Estimating Moral Hazard and Risk-Selection Effects

For policy, cost savings in Swiss Francs (CHF) rather than in percent are of interest. Unlike relative savings, these strongly depend on the expenditure level of the subpopulation who chooses the respective contract. The results are displayed in Table 3.5. For estimate (1), expected HCE according to type of contract is estimated by predicting the probability of positive HCE times the amount of HCE. For instance, individuals with the baseline contract had expected HCE of CHF 4,320 (the reference value), while those with a high deductible combined with the IPA option had CHF 1,100 only. These values are derived from the full specifications displayed in Table 3.4, which control for both health-related and other determinants of contract choice.

Note that the value of CHF 1,100 is the estimated average expenditure of the individuals *who actually chose the high-deductible contract.* In order to estimate the expenditure of the *same* subpopulation assuming they had chosen the baseline contract, the dummies for deductibles (and IPA plans, respectively) are set to zero when predicting expected HCE. The results are shown as estimate (2) of Table 3.5. Since both estimates (1) and (2) pertain to the same subpopulation of individuals, their difference represents the influence of moral hazard.

Table 3.5: Estimated Cost Reductions in Swiss Francs, 2006.

	$\widehat{HCE}$ *according to subpopulations:* Actual choice of contract (1)	Simulated baseline contract (2)	Moral hazard (2) − (1)	Risk selection (see text)
A. FFS plans				
Minimum DED (baseline)	4,320			
	(29)			
Medium DED	3,180	3,430	250	891
	(37)	(34)		
High DED	1,100	1,422	322	2,898
	(23)	(57)		
B. IPA plans				
Minimum DED	2,985	3,340	355	980
	(99)	(133)		
Medium DED	2,023	2,374	351	1,946
	(159)	(124)		
High DED	843	1,111	268	3,209
	(55)	(78)		

Standard errors in parentheses, DED = Deductible, IPA = Independent Practice Association.

For instance, the effects of demand-side cost sharing can be deduced from estimate (2) for the high-deductibles subpopulation in Panel A of Table 3.5. This is the same subpopulation that gives rise to estimate (1), with determinants of contract choice held constant. The only difference is that their predicted HCE is derived by setting the DED dummy equal to zero. Since treating physicians were confronted with the same incentives, the difference of CHF 322 (= 1,422 − 1,100) in all likelihood is caused by the difference in demand-side incentives.

This estimate can now be compared to the effect of supply-side incentives. In Panel B of Table 3.5, estimate (1) for the subpopulation in the minimum-deductible category amounts to CHF 2,985, while estimate (2) amounts to CHF 3,340. Again, both estimates refer to the same subpopulation, except that the IPA dummy is set to zero in estimates (2). Since determinants of contract choice are controlled for and the same minimum deductible is applied to the subpopulation, the difference of CHF 355 (= 3,340 − 2,985) can be attributed to the difference in supply-side incentives.

When comparing the entries of Panels A and B of Table 3.5, one is led to conclude that the effects of supply-side and demand-side cost sharing have about the same magnitude. However, both types of moral hazard mitigation come at the price of considerable risk-selection effects. They can be estimated as follows. In Panel A of Table 3.5, the high-deductible subpopulation has an estimated HCE of 1,100, while the subpopulation with the baseline (minimum-deductible) contract has CHF 4,320. With the determinants of contract choice not held constant this time, the difference of CHF 3,220 is caused by both moral-hazard and risk-selection effects. Mitigation of demand-side moral hazard has been estimated at CHF 322 above; therefore, the remainder of CHF 2,898 (3,220 − 322) needs to be attributed to risk selection. Turning to supply-side cost sharing in Panel B of Table 3.5, the high-deductible IPA subpopulation is seen to have estimated HCE of CHF 843, even CHF 3,477 below the benchmark value of CHF 4,320. However, only CHF 268 of this difference can be traced to an attenuation of moral hazard, leaving CHF 3,209 (3,477 − 268) as the likely effect of risk selection. A comparison of the entries of Panels A and B of Table 3.5 reveals that supply-side cost sharing in combination with demand-side cost sharing seems to go along with even more marked risk-selection selection effects than demand-side cost sharing combined with FFS.

Overall, the estimates of Table 3.5 point to a conflict of interest. If the objective is simply to reduce HCE, high-deductible plans, preferably combined with the IPA option, fare best. If however the objective is to achieve a favorable balance between moral hazard reduction and risk-selection effects, the minimum-deductible IPA option ranks first, with a ratio of 355/980, or 1 : 2.76, followed by the medium-deductible FFS alternative with 250/891 or 1 : 3.56. These ratios are of interest to policy makers or regulators who wish to introduce voluntary cost-sharing plans in order to mitigate moral hazard, but at the same time wish to avoid risk-selection effects that lead to market segmentation and hence high premiums for high risks. If the risk-selection problem is regarded as severe, plans with low ratios of moral hazard reduction versus risk selection are preferable.

3.6.3 Tests of Validity

Two types of validity tests are performed in this section. A first set contains modifications of the two-part model in terms of observation period, estimation of the health status proxy, and economic specification. The second type of test focuses on individuals who switched away from the minimum-deductible plans between 2003 and 2006.

In the two columns (A) of Table 3.6, results are shown for a pooled GLM estimation over the years 2004 - 2006 rather than just 2006 (standard errors are clustered by patient). In an attempt to measure variations in health status over time, the proxy now is calculated per year using data from the respective previous year. Comparison with the lower part of Table 3.4 shows the estimated incentive effects to be stronger. For instance, the coefficient pertaining to the medium deductible category is -0.137 here, but only -0.042 in Table 3.4. The likely reason is that the incidence of chronic illness and hence risk selection effects are controlled for to a lesser degree because only one year of data is used for the calculation of the health proxy in test A.

Columns B of Table 3.6 show the results of pooled estimations over the years 2005 and 2006 only. Here, the health proxy was calculated over two rather than the three previous years prior to averaging. Again, the estimated moral hazard effects tend to exceed the ones of the original model, presumably because the two-year proxy is less effective in controlling for chronic illness than the three-year proxy.

Columns C of Table 3.6 address the fact that prior HCE is influenced by incentives. As a consequence, a proxy based on prior HCE may make individuals with high deductibles or managed-care type plans appear healthier than they are. To gauge the extent of this potential bias, we recalculated the health proxy by augmenting the observed HCE of individuals in both types of cost-sharing plans, using the estimated coefficients of the incentive effects (lower part of Table 3.4, full model). As expected, the resulting moral hazard effects are stronger than in Table 3.4, but not dramatically so.

Table 3.6: Specification Tests.[a]

	P(HCE>0), GLM with probit link			HCE|HCE>0, GLM with log link			FE
Test	A	B	C	A	B	C	D
Data for 2PM	2004-06	2005-06	2006	2004-06	2005-06	2006	2004-06
Medium DED	-0.137***	-0.097***	-0.058***	-0.117***	-0.074***	-0.091***	-0.003
	(0.007)	(0.009)	(0.015)	(0.011)	(0.015)	(0.016)	(0.014)
Med. DED * IPA	0.055	0.080	0.093	-0.037	-0.047	0.020	-0.154*
	(0.043)	(0.059)	(0.062)	(0.060)	(0.081)	(0.079)	(0.071)
High DED	-0.441***	-0.329***	-0.233***	-0.265***	-0.185***	-0.218***	-0.149***
	(0.008)	(0.011)	(0.025)	(0.016)	(0.021)	(0.033)	(0.017)
High DED * IPA	0.093***	0.070*	0.062	-0.041	-0.077	0.090	-0.091
	(0.024)	(0.030)	(0.038)	(0.048)	(0.056)	(0.060)	(0.049)
IPA	-0.100***	-0.061*	0.053	-0.284***	-0.201***	-0.145*	-0.108*
	(0.022)	(0.029)	(0.044)	(0.048)	(0.060)	(0.059)	(0.043)
N	475,107	325,442	163,686	372,091	256,166	128,744	372,091

[a] See text for explanation of test A through D. *p<0.05, **p<0.01, ***p<0.001, Standard errors in parentheses. Additional regressors are the same as those in Table 3.4.

Finally, a fixed effects specification was used for the second part of the two-part model (column D of Table 3.4). The dependent variable here is the log of HCE over the years 2004 - 2006. For the high-deductible plan and the IPA, the estimated incentive effects are close to those in Table 3.4. For the medium deductible, the direct incentive effect is weaker and insignificant, but the interaction term with the IPA is stronger.

Still another possibility to test the validity of the results presented in Section 3.6.2 is to track the HCE of individuals who switch between contract types. To an approximation, their personal characteristics are unchanged while contractual incentives are modified.[11] Therefore, it is of interest to compare the difference in HCE prior and after the change, covering the years 2003/04, 2004/05, and 2005/06 with the moral hazard effects displayed in Table 3.5. For simplicity, only switches away from the baseline (minimum-deductible, FFS) contract are retained (see Table 3.7). The change in HCE turns out to be symmetrically distributed, permitting estimation of an untransformed linear random effects model. Attribution of annual HCE to contracts hardly causes problems because switches usually take place at the beginning

[11] This statement is only approximately true because the determinants of contract choice must have changed.

of the year. In order to be able to use several years of data, the health proxy is calculated using the respective previous year only rather than averaging over three years.

Table 3.7: Analysis of Switchers, 2003/04, 2004/05, 2005/06.

Switch from baseline contract	Numbers of switchers	Random effects coefficient	SE
A. FFS plans			
... to medium DED	2,139	-212.99	(130.06)
... to high DED	13,503	-387.99***	(53.27)
B. IPA plans			
... to minimum DED	1,905	-260.27	(111.07)
... to medium DED	88	-189.23	(454.35)
... to high DED	805	-219.12	(161.17)

*p<0.05, **p<0.01,***p<0.001, Standard errors in parentheses. Additional regressors are the same as those in Table 3.4.

For the switchers from the baseline to a medium-deductible FFS contract, the estimated HCE reduction amounted to CHF 213, which is similar to the CHF 250 of Table 3.5 (Panel A), indicating the attenuation of moral hazard relative to the baseline contract. However, the standard error is large due to the small number of switchers in combination with substantial year-to-year variations in HCE. The switch from the baseline to a high-deductible contract is estimated to be associated with a reduction in HCE amounting to CHF 388, which is again comparable to the figure from Table 3.5 (CHF 322). This group contains over 13,000 individuals, resulting in a lower standard error and statistical significance.

In the case of supply-side cost sharing, switches from the baseline FFS contract to an IPA option are associated with estimated cost reductions that are again compatible with those evidenced in Table 3.5. As to the one apparent exception in Table 3.4 (transition to the medium-deductible, IPA contract), the number of switchers is too low to permit statistical inference.

The evidence compiled in this section comes from two sources. The first consists of several modifications in the estimation of the two-part model. The second source

is the analysis of changes in HCE that go along with switching contracts. On the whole, none of the validity tests performed suggests that the effects of supply-side and demand-side cost sharing presented in Table 3.5 are a mere chance result.

3.6.4 Which Types of Medical Care Are Most Affected by Cost Sharing?

The encouraging outcome of the validity tests presented in Section 3.6.3 motivates a more detailed analysis of the effects of demand-side and supply-side cost sharing. The estimation technique described in Section 3.5 (with the specification in keeping with the full model of Table 3.4) is applied to general practitioners' services, specialists, drugs, physical therapy, outpatient hospital services, and inpatient hospital services. Therefore, the estimated coefficients reported in Table 3.8 reflect incentive effects of the two types of cost sharing, with selection effects controlled for (coefficients of the control variables are not shown). For each cost-sharing option, the first line pertains to the first part of the two-part model and second line to the second part, respectively. The interaction terms of deductibles and IPA plans are not shown for brevity. In line with the four-part model advocated by Duan et al. (1982), the probability of observing positive hospital inpatient expenditure is estimated only for the individuals with positive ambulatory care expenditure.

As reported in Table 3.8, demand-side cost sharing in the guise of a medium deductible significantly reduces the probability of expenditures on GP services, drugs, and hospital outpatient services. For specialized medicine, a significantly positive influence is estimated, which contradicts intuition. As to HCE given that it is positive, there are consistent indications of a reduction effect, which however attains statistical significance in the case of GP services and drugs only.

A high deductible does go along with a decreased probability of all types of care. Moreover, its estimated reduction effect consistently exceeds that of a medium deductible. The impact on hospital inpatient care is surprising as most patients already have HCE in excess of the deductible when entering a hospital. However, patients

Table 3.8: Estimation Results from the Two-part Model According to Type of Care.

		GP	Specialist	Drugs	Physical Therapy	Hospital Outpatient	Hospital Inpatient
Med. DED	P(HCE>0)	-0.034***	0.024*	-0.076***	0.004	-0.019*	-0.005
		(0.010)	(0.009)	(0.012)	(0.010)	(0.009)	(0.012)
	HCE|HCE>0	-0.048***	-0.018	-0.090***	-0.026	-0.045	-0.045
		(0.008)	(0.013)	(0.020)	(0.015)	(0.026)	(0.026)
High DED	P(HCE>0)	-0.270***	-0.143***	-0.340***	-0.107***	-0.092***	-0.039**
		(0.019)	(0.019)	(0.023)	(0.024)	(0.020)	(0.015)
	HCE|HCE>0	-0.154***	-0.087***	-0.137***	0.004	0.039	-0.050
		(0.011)	(0.018)	(0.038)	(0.021)	(0.039)	(0.036)
IPA	P(HCE>0)	0.088***	0.103**	0.087*	-0.049**	-0.058***	-0.066**
		(0.016)	(0.032)	(0.037)	(0.018)	(0.016)	(0.022)
	HCE|HCE>0	-0.084***	-0.153***	-0.156**	-0.058*	-0.088	-0.116*
		(0.016)	(0.023)	(0.053)	(0.029)	(0.046)	(0.046)
AIC:	P(HCE>0)	1.042	1.109	0.822	0.820	1.061	0.614
	HCE|HCE>0	14.41	15.31	15.233	15.254	16.043	19.751
N:	P(HCE>0)	163,686	163,686	163,686	163,686	163,686	128,744
	HCE|HCE>0	101,265	86,208	112,941	28,241	46,923	17,205

*p<0.05, **p<0.01, ***p<0.001, standard errors in parentheses. Additional regressors are the same as those in Table 3.4.

with high deductibles are less likely to initiate the whole process of diagnostic testing and procedures which ultimately may result in hospitalization [Zweifel (1992) found similar results using German data]. Given positive HCE, there is clear evidence of moral hazard attenuation for GP services, specialist services, and drugs which again exceeds the amount found for the medium deductible.

Turning to supply-side cost sharing, IPA plans exhibit an increased probability of use of GP services, specialist services, and drugs but a decreased probability in the case of physical therapy and hospital services, as expected for the latter. As to the second part of the two-part model, the IPA options are associated with a reduction of expenditure on all types of HCE, with the only exception of hospital outpatient services. As fees of specialists and hospitals are regulated to be equal for FFS and IPA contracts, these effects are exclusively due to a reduction in quantity.[12]

[12]In order to validate our results, we reestimated the second part of the two-part model by OLS on log expenditure. The estimates are close and equal in sign to those in Table 3.8. The only exception is the coefficient for hospital outpatient services, which is significantly negative in the OLS estimation. It is not a priori clear which estimate is more plausible for this heterogenous patient group. Some patients are chronically ill and in need of repeated procedures (suggesting no effect), while others visit the emergency room for relatively minor ailments (where deductibles might well be effective).

Summing up, the evidence of Table 3.8 suggests that supply-side and demand-side cost sharing are effective in different ways. On the demand side, high deductibles primarily seem to lower the likelihood of seeing a GP or a specialist and of using drugs. They also markedly reduce expenditure on GP services, but have no effect on hospital inpatient expenditure. By way of contrast, supply-side cost sharing in the guise of IPA plans is even associated with an increased probability of calling on services of GPs and specialists and of using drugs, presumably due to increased use of preventive care. Its expenditure-reducing effect is concentrated on specialists and inpatient hospital services.

These findings can be compared to the famous RAND Health Insurance Experiment (Manning et al. (1987)). The demand-side cost sharing plans in the HIE required the patient to pay a percentage of care out of pocket, up to a stop-loss. In line with the results in Table 3.8, these plans reduced the probability of incurring any medical expenses, the amount of outpatient expenses, and the probability of inpatient expenses. They did not significantly alter the expenditure for inpatient services.[13] The HMO plan in the HIE was a prepaid group practice. Compared to a FFS plan with no demand-side cost sharing, the HMO plan had a markedly lower hospitalization rate. The effects on hospitalizations are stronger than those measured by this study, which is likely due to stronger financial incentives in the HMO analyzed by RAND, than the IPA analyzed here.

3.7 Discussion

The aim of this section is to discuss the policy implications of our results, relating them to recent literature. A first salient point is that the estimated absolute cost reduction of CHF 250 (see Panel A of Table 3.5) due to a deductible of CHF 500 rather than 300 exceeds the maximum increase in out-of-pocket expenditure (CHF

[13]The HIE also included a deductible plan. However, this plan is not directly comparable to the Swiss case because it only applied to outpatient services. It reduced outpatient expenditure and the probability of medical care, but not the probability or amount of inpatient care.

200 = 500 − 300). This is confirmed by two other recent studies using Swiss data from earlier years when the minimum deductible was CHF 230 and the next lowest, CHF 400. They both seek to control for risk selection effects. Van Kleef et al. (2008) estimate that raising the deductible from CHF 230 to 400 (i.e. by CHF 170) would serve to reduce expected HCE by CHF 382 (see Table 3.9). Gardiol et al. (2005) take the maximum deductible of CHF 1,500 as their reference point, calculating the incentive effects from there. The transition from the medium deductible of CHF 400 to the minimum of CHF 230 is estimated to generate 'true' savings of CHF 185 (= 697 − 512), which again exceeds the out-of-pocket difference of CHF 170. A possible explanation of this 'overshooting' is that patients, who are usually not well informed about the cost of medical care, do not know when they exceed the deductible.

Table 3.9: Estimated Incentive Effects of Demand-side Cost Sharing on HCE, in CHF.

Deductible levels	230	400	600	1,200	1,500
Effects reported by Van Kleef et al. (2008)	-3	-382	-443	-276	-318
Effects reported by Gardiol et al. (2005)	+697	+512	+306	+62	0
Deductible levels	300	500	1,000 to 2,500		
Effects reported by this study		-250	-322		

CHF 1≈ EUR 0.66

The second point relates to risk adjustment (RA). Note from Table 3.5 that estimated moral hazard reductions not only fall far short of gross differences in expected HCE as indicated by estimates (1) but are markedly plan-specific. As noted by Van Kleef et al. (2008) and Van Kleef et al. (2006), this varying mix of risk-selection and moral hazard effects poses a great challenge to regulators in a system combining community rating with RA. The issue is the extent to which insurers should be allowed to pass on gross savings to consumers. The appropriate amount seems to be the amount of 'true' savings net of risk-selection effects. Yet, Van Kleef et al. (2006) show that if only very low risks opt for higher deductibles at first, premium reductions reflecting 'true' savings are too small to create incentives for choosing these options.[14] As a remedy, they propose not to entirely net out risk-

[14]This reflects the Swiss experience after the introduction of voluntary deductibles in 1996.

selection effects for determining allowable premium reductions. Empirical evidence by Van Kleef et al. (2008) reveals that the current RA schemes of the Netherlands and Switzerland do leave room for risk-selection effects in premium reductions. Their finding is replicated by this study since the maximum allowable premium reductions for high deductibles range between CHF 560 and 1,760 (see Table 3.1), exceeding by far the CHF 322 and 268, respectively that can be attributed to the attenuation of moral hazard (see Table 3.5).

3.8 Conclusions

Managed competition in social health insurance aims at creating incentives for insurers to increase efficiency and respond to consumer preferences while preserving solidarity between high- and low-risk types [Van de Ven et al. (2007)]. Therefore, it is important to know whether contractual innovations such as deductibles or capitated IPA plans achieve 'true' cost savings rather than merely serving as a means for risk selection. This research measures and compares the impacts of demand-side cost sharing (through voluntary deductibles) and supply-side cost sharing (through prepaid IPA plans) on individual health care expenditure (HCE), controlling for risk-selection effects. The data comes from a large panel of Swiss adults covering the years 2003 to 2006. Since unobserved health status influences both contract choice and HCE, a proxy is constructed from HCE during the first three years of the observation period, complemented by the residuals from the contract choice equation [the two-stage residual inclusion method proposed by Terza et al. (2008)].

Higher annual deductibles and IPA plans are both found to achieve marked reductions of moral hazard. An increase in the annual deductible by CHF 200 (some EUR 133, from minimum to medium) is estimated to decrease the probability of positive HCE by almost 1 percentage point, while the IPA alternative might even be associated with an increase. In return, it achieves a reduction of positive HCE by some 12 percent, compared to only 7 percent of the medium deductible. Increasing

the deductible by CHF 700 (some EUR 466) reduces the probability of reporting HCE by about 3.7 percent and the amount of positive HCE by about 18 percent.

However, this effectiveness of demand-side cost sharing comes at the price of substantial risk-selection effects. Because voluntary cost sharing plans are especially attractive to low risks, such plans might lead to market segmentation and hence higher premiums for high risks. The most favorable ratio of moral hazard attenuation over risk selection is achieved by the IPA with the minimum deductible, amounting to 1 : 2.76 (CHF 355 / 980; HCE amounts to CHF 4,610 for the baseline contract). The next-best alternative is the medium-deductible FFS contract with a ratio of 1 : 3.56 (CHF 250 / 891). According to this criterion, supply-side cost sharing is somewhat more effective than the demand-side alternative.

Still, this research is subject to several limitations. First, since the data set only comprises individuals who were with one and the same insurer from 2003 to 2006, it fails to measure risk-selection effects associated with changes between competing insurers. Second, even 'within' risk-selection effects may not be controlled for perfectly. There is no guarantee that the HCE equation is correctly specified for the three preceding years, a necessary condition for obtaining residuals that serve as good proxies for unobserved health. The same caveat applies to the residuals of the contract choice equations. Thus, estimates of expected HCE reductions achieved by higher deductibles and IPA plans could still be biased. Third, deductible options have price and income effects. As shown by Nyman (1999), only the former should be counted as inefficient consumption. Finally, results relating to IPA plans have limited generality as long as they cannot be linked in detail to the incentives faced by participating health care providers.

Nevertheless, the findings of this study permit one to draw the conclusion that allowing insurers to offer plans with both demand-side and supply-side cost sharing does generate 'true' savings in Swiss social health insurance. After controlling for risk-selection effects, both variants are estimated to achieve marked reductions in

moral hazard that can be passed on to consumers in the guise of premium reductions without jeopardizing insurers' solvency.

Acknowledgments

The authors would like to thank the CSS Institute for Empirical Health Economics for providing data, financial support, and many insights. We are grateful to three anonymous reviewers as well as Randall Ellis, Wynand Van de Ven, Hansjoerg Lehmann, Johannes Schoder, Stefan Pichler, Lorens Helmchen, Matthias Kifmann, Ute Kunze, Ute Studer-Merkle, Bernhard Keller, Stefan von Rotz, Urs Kaeser, and participants at the 2011 Darmstadt Conference on Inequality, Insurance and Health for helpful comments and discussion.

3.9 Appendix

3.9.1 Estimating a Proxy for Health Status

Table 3.10: Estimation of Equations (3.1) and (3.2).

	P(HCE_{it} >0) Random Effects Probit	$HCE_{it}\|HCE_{it}$ >0 Random Effects AR (1)
Greater Metropolitan Area	0.018	-0.077***
	(0.021)	(0.010)
Affluent Community	0.101*	-0.069***
	(0.041)	(0.020)
Regional Center	-0.083**	-0.203***
	(0.032)	(0.017)
Rural, mainly industrial	-0.070**	-0.173***
	(0.025)	(0.013)
Rural, agriculture	-0.171***	-0.209***
	(0.028)	(0.015)
Berne city	0.088**	0.037*
	(0.033)	(0.016)
Lucerne city	-0.059*	-0.241***
	(0.028)	(0.014)
Geneva city	0.396***	0.385***
	(0.042)	(0.019)
2004	0.007	0.091***
	(0.012)	(0.005)
2005	0.017	0.089***
	(0.012)	(0.005)
Constant	1.928***	6.911***
	(0.036)	(0.018)
N	253,653	218,208
α	.691	.531
ρ		.090

*p<0.05, **p<0.01, ***p<0.001, Standard errors in parentheses. Additional regressors are age, gender, and additional region-specific dummies.

α : Fraction of error variance due to individual-specific term.

ρ : Estimated autocorrelation coefficient.

3.9.2 Predicting Contract Choice

Table 3.11: Estimation of Contract Choice in 2006.

	Choice of Deductible Ordered Probit	Choice of IPA Probit
Health proxy	-0.341***	-0.219***
	(0.006)	(0.012)
(Health proxy)^2	0.042***	0.055***
	(0.003)	(0.005)
(Health proxy)^3	-0.002***	-0.004***
	(0.000)	(0.001)
Bad credit record	-0.274***	-0.359***
	(0.014)	(0.027)
Years of CSS membership since 1999	-0.042***	-0.052***
	(0.004)	0.007)
Baseline premium	0.018***	0.005***
	(0.002)	(0.001)
Premium reduction for medium DED	0.008	
	(0.010)	
Premium reduction for high DED	0.015***	
	(0.003)	
Premium reduction for IPA		0.013***
		(0.001)
IPA operational in zipcode area		1.436***
		(0.058)
Constant		-3.926***
		(0.287)
Cut points	7.145 / 7.823	
Log likelihood	-140,618	-27,297
Number of observations	163,686	163,686

*p<0.05, **p<0.01, ***p<0.001, Standard errors in parentheses. Additional regressors are age, gender, types of municipalities, region specific dummies, accident coverage, long term care coverage, and youth rebate eligibility.

Generic Substitution, Financial Interests, and Imperfect Agency

Maurus Rischatsch, Maria Trottmann, Peter Zweifel

Abstract: Policy makers around the world seek to encourage generic substitution. In this paper, the importance of prescribing physicians' imperfect agency is tested using the fact that some Swiss jurisdictions allow physicians to dispense drugs on their own account (physician dispensing, PD) while others disallow it. We estimate a model of physician drug choice with the help of drug claim data, finding a significant positive association between PD and the use of generics. While this points to imperfect agency, generics are prescribed more often to patients with high copayments or low incomes.

Keywords: Physician agency, prescribing behavior, drug dispensing, generic substitution, brand-name drugs

JEL classification: I10; I18; I19

Chapter 4

Generic Substitution, Financial Interests, and Imperfect Agency

4.1 Introduction

Policy makers around the world seek to encourage generic substitution (i.e. the replacement of brand-name by generic drugs) in an attempt to reduce the pharmaceutical bill. In the United States for instance, several state policies promote the use of generic products by Medicaid beneficiaries [CMS (2004)]. Similar initiatives exist in Germany [Leutgeb et al. (2009)], Sweden [Andersson et al. (2007)], Switzerland [Decollogny and Ruggli (2006)], and Japan [Matsuda (2008)]. To be successful, these initiatives must be aligned with prescribing physicians' (or pharmacists') incentives. Generic substitution not only requires effort and time on the part of these professionals but also entails the risk of meeting with patient resistance. Three components of prescribers' utility can work to overcome resistance against generic substitution. First, prescribers may earn higher contributions to income from generic than from brand-named drugs. Second, acting as agents by taking patients' total (rather than merely health-related) utility into account, physicians are predicted to prescribe the generic if the savings accruing to the patient are important enough. Third, in view of public concern about growing health care expenditure, cost savings accruing to insurers might motivate physicians to prescribe lower-priced generic drugs.

In this context, evidence from Switzerland is of considerable interest. In some Swiss jurisdictions (cantons), physicians are allowed to dispense drugs to their patients on their own account. This setting will be referred to as 'physician dispensing' (PD) in the remainder of this paper.[1] In the remaining jurisdictions, physicians are obliged to let a pharmacy fill their prescriptions. Thus, both the PD and the non-PD (i.e. pharmacy-based) setting can be observed under otherwise very similar conditions. PD may well affect generic substitution provided physicians act as imperfect agents and given that generic drugs differ from brand-name drugs in terms of their contribution to physician income.

Retail prices paid by patients are regulated to be equal for all drug sellers (physicians and pharmacies). The contribution to the sellers' income, then, is the difference between manufacturers' prices and retail prices. Concerning manufacturers' prices, there is room for discounts and individual bargaining, causing the effective contributions to income to be unknown. However, several factors indicate that contributions to physician income can be higher for generic than for brand-name drugs. First, many generic alternatives are usually available for the same brand-name drug, leading to fierce competition for access to prescribers among generic producers. Second, the retail prices of generic drugs are markedly higher in Switzerland than in comparable European countries, suggesting that generic producers have ample leeway for rebates to prescribers.[2] Third, while there is no public information about such rebates, interviews conducted with Swiss wholesalers and physicians support the notion that prescribers derive more income from generic than brand-name drugs.

The remainder of this article is structured as follows. Section 4.2 contains a short review of the literature. Section 4.3 describes the institutional setting. Section 4.4 presents a theoretical model of physician prescribing behavior, along with a set of testable hypotheses. The empirical strategy used for hypothesis testing is explained

[1] PD is the counterpart of prescribing pharmacists, who exist e.g. in the case of refills in the United States, Canada, the United Kingdom, and New Zealand [Emmerton et al. (2005)]. In both cases, the prescriber and the dispenser is one and the same person or institution, respectively.

[2] The prices for brand-name drugs are also higher in Switzerland, but the markups for physicians are smaller (see Section 4.3.2).

in Section 4.5. Section 4.6 contains a description of the data. Results are shown in Section 4.7, while Section 4.8 rounds off with a summary and conclusions.

4.2 Literature Review

To keep this survey concise, there will be no discussion of research into physician behavior in general. Rather, focus is on prescribing behavior. An early pertinent study is the one by Morton-Jones and Pringle (1993), who compare prescription patterns of PD and non-PD providers in the UK, finding that the share of generic drugs is lower in the PD segment. Liu et al. (2009) analyze the choice between generic and brand-name drugs in Taiwan, where PD is the dominant mode. According to them, financial incentives markedly influence this choice. Specifically, providers on a global budget are more likely to prescribe generic drugs than those reimbursed fee-for-service. Moreover, cheaper brand-name drugs (which in Taiwan contribute less to physician income, as in Switzerland) are more often replaced by generics than expensive ones. Using Japanese data on hypertension drug sales, Iizuka (2007) concludes that markups available to physicians significantly influence drug choice. However, he also finds that physicians take the cost of the drug to their patients into account. Finally, the 2000 reform in South Korea provides an interesting natural experiment. At that time, both drug dispensing by physicians and drug prescribing by independent pharmacists were outlawed. Descriptive statistics presented by Kim and Ruger (2008) indicate a marked increase in the market share of high-price drugs in the year following the reform. However, the longer-term effects of the reform could not be assessed on the basis of their data.

Papers that are methodically related to ours are Hellerstein (1998), Lundin (2000), and Hellstrom and Rudholm (2010). They analyze the choice between generic and brand-name drugs in a non-PD setting. Hellerstein argues that physicians bear higher information costs when prescribing generic rather than brand-name drugs because they have more personal experience with the brand-name than with the generic drugs. Contrary to the hypothesis of perfect agency, she finds that prescription is not influ-

enced by patients' insurance status and hence financial burden. However, physicians who predominately treat patients in capitated or Health Maintenance Organization (HMO) settings are more likely to prescribe generics (controlling for individual insurance status). Her panel data specification also shows that a large part of the unexplained variance is physician-specific, which also holds true of Lundin's contribution. Interestingly, Lundin argues that physicians may want to honor R&D expenditure and pioneering effort by innovators, causing them to bear added psychic cost when prescribing a generic. He finds evidence that higher cost to the patient through copayment increases the probability of generics being prescribed, while higher cost to the insurer does not. Hellstrom and Rudholm argue that the uncertainty about the quality of generic drugs incites physicians to prescribe brand-name drugs. Their empirical evidence shows that physicians are indeed less likely to allow generic substitution for older (and presumably sicker) patients. However, their measure of uncertainty about quality came out insignificant in the decision equation.

Another reason why the prescription of generic drugs might require extra effort on the part of the physician is given by Griliches and Cockburn (1994). They argue that many patients perceive generic drugs as less safe and of lower quality, making the patient suffer a 'putative loss' when using them. Therefore, a physician prescribing the generic drug needs to convince the patient of its bioequivalence.

To our knowledge, there is no Swiss study that analyzes the effect of PD on the choice between generic and brand-name drugs. The one exception is Hunkeler (2008) who presents corroborating evidence for the hypothesis that PD leads to margin optimization or even margin maximization[3] through dispensing packages and dosages with higher official physician margins. These packages are launched first by companies entering the generics market; later, they are complemented by additional package sizes and dosages (for more institutional detail regarding Swiss health insurance, see Section 4.3). The other studies of PD in Switzerland have focused on its impact

[3]The difference between margin optimization and maximization is that in the first case, PD providers prescribe several small packages instead of one large package while in the second case, they prescribe a higher quantity to maximize their income.

on total physician billings or health care expenditure (HCE), respectively. An early investigation by Zweifel (1985) concluded that while PD creates incentives to keep patients out of the hospital (where different physicians are in charge as a rule), the savings achieved through a reduced rate of hospitalization fall short of the extra drug expenditure induced in ambulatory care. At a more aggregate level, Dummermuth (1993) compares two otherwise similar neighboring cantons (Lucerne with PD and Argovia without PD), finding PD to be associated with slightly higher per capita drug expenditure as well as HCE. This finding is in line with Beck et al. (2004), who relate per-capita drug expenditure to several properties of cantons, among them, their PD status. By way of contrast, Vatter and Ruefli (2003), who control for a very comprehensive set of political and socioeconomic covariates, indentify a significantly negative effect of the share of PD providers on per capita HCE. More surprisingly still, Schleiniger et al. (2007) estimate a significantly negative effect of PD on cantonal drug expenditure which is robust across several specifications.

4.3 Institutional Setting

Basic health insurance coverage in Switzerland written by some 80 competing private not-for-profit insurers is mandatory for a broad basket of services and drugs. Physicians in private practice are mostly paid according to a nationwide uniform fee schedule called TARMED [see Zweifel and Tai-Seale (2009) for description and criticism].[4] Provision of health care is decentralized and the 26 Swiss cantons ('jurisdictions') have considerable say in its regulation, including the regulation of drug dispensing.

4.3.1 Physicians' Dispensing Rights

Thirteen of the twenty-six Swiss cantons give dispensing rights to all physicians, seven apply mixed systems while six generally disallow PD. Physicians who dispense on

[4] A small number of physicians works in managed-care type arrangements, where other modes of payment are possible.

average derive about 18 percent of their revenue from PD. This number is higher for general practitioners (28 percent) and lower for specialists (8 percent) [see Hunkeler (2008)]. Therefore, the financial incentives linked with the amount and structure of PD are substantial. Acknowledging problems of asymmetric information between physicians and patients, some cantons with PD require physicians to inform patients about their right to obtain a prescription to be filled by the pharmacy of their choice.

In the context of the present study, an important question is whether cantons that allow PD attract substantially different types of physicians than do non-PD cantons. Since the data is provided by a health insurer, they do not contain information about the determinants of locational choice such as regional origin of the physician and her spouse, or the location of her medical school. This makes an analysis of physicians' choice of location impossible. Moreover, it is known that young physicians mainly take over existing practices rather than opening new ones in response to large administrative hurdles, pointing to a narrowed choice of location. Still, if physicians who are very susceptible to financial incentives are disproportionately located in the PD cantons, our estimates in Section 4.7 might be upwardly biased.[5]

4.3.2 Contributions to Income from Drug Dispensing

For non-PD practitioners, the contribution to income from dispensing is zero. For PD practitioners, the contribution earned by selling a specific drug consists of three components, namely (i) a fixed lump sum, (ii) a percentage of the regulated manufacturer price, and (iii) discounts that are conceded to physicians by pharmaceutical companies. The first two components are regulated by the government and published in official registers. The third component is the outcome of an individual bargaining process between prescriber and sales representative, which is unobservable to us. However, they ultimately reflect the bargaining position of the pharmaceutical company, about which a few facts are known.

[5]This may be true although dummy variables for cantons and community types are included in the estimation in order to control for differences between regions (see Section 4.7).

According to Liu et al. (2009), the discount on manufacturers' prices offered increases with market size, competition, and retail price but decreases with marginal cost. First, market size is small in Switzerland for both brand-name and generic drugs. With regard to competition, the market usually contains one brand-name drug only but a large number of generic alternatives (more than 10 in this analysis). Therefore, producers of generic drugs are more likely to use discounts in their attempt to increase market share. Next, marginal cost of brand-name and generic drugs can be assumed equal in the present setting.

In addition, international comparisons of reimbursement prices offer indirect evidence suggesting that generic producers in Switzerland have ample leeway for discounts. For fixing the reimbursement price of brand-name drugs, Switzerland uses a reference group comprising Germany, Denmark, UK, the Netherlands, France, Italy, and Austria. Reimbursement prices for generic drugs have to be at least 40 percent lower than those of the original drug. However, this does not imply that generic producers earn lower effective margins. In fact, Santesuisse (2009) and IMS (2009) calculate price indexes for drugs with and without patent protection for Switzerland and the seven countries cited above. The two studies conclude that both prices for band-name (p_b) and generic drugs (p_g) are higher in Switzerland, i.e. $\Delta p_b = p_b - p_b^R > 0$ and $\Delta p_g = p_g - p_g^R > 0$, where R denotes the average drug price in the reference group. But they also find that the international price difference is larger in the case of generic than for brand-name drugs ($\Delta p_g > \Delta p_b$). Assuming that producers have the same cost structure in Switzerland and elsewhere, the extra profit margin earned in Switzerland is therefore higher for generic than for brand-name producers, i.e. $\tilde{m} = \Delta p_g - \Delta p_b > 0$. They can use their net advantage $\tilde{m}$ for inducing physicians to prescribe their products.

In all, manufacturers of generic drugs are likely to offer larger discounts to physicians than brand-name producers. Indeed, interviews conducted with Swiss wholesalers and physicians support the notion that prescribers derive more income from

generic drugs, although no market participant is willing to publish the exact discounts that are offered or accepted.

In the context of the present study, it is important to note that the law forbids to give, promise or accept any monetary or monetary equivalent reward for the prescription of a specific drug. Therefore, manufacturers are not allowed to promise rewards (for example higher discounts) for the achievement of a higher sales volume.

4.3.3 Copayment Arrangements

Prescription drugs are covered by compulsory health insurance, which kicks in when the annual deductible is exceeded. The minimum annual deductible amounts to CHF 300 ($\approx$ US$ 250 at 2007 exchange rates). Voluntary deductibles range from CHF 500 to 2,500 (US$ 415 to 2,100) and are chosen by the insured at the beginning of the year. The deductible applies to all health care services except those related to maternity. When the deductible is exceeded, there is a 10 percent rate of coinsurance up to a stop-loss of CHF 700 (US$ 580) per year. For instance, a patient with a deductible of CHF 2,500 would spend a maximum of CHF 3,200 (US$ 2,700) out of pocket. For certain brand-name drugs, the rate of coinsurance was increased to 20 percent during our observation period (2005 to 2007). However, producers of brand-name drugs can escape this increased coinsurance by lowering their prices. As a consequence of different deductibles and changing rates of coinsurance, some patients have a stronger interest in receiving cheaper drugs than others.

4.4 Theoretical Model of Physicians' Drug Choice

Because of their central role in the resource allocation in health care markets, the behavior of physicians has spawned a very rich literature (see McGuire (2000) for an overview). The purpose of this section is to derive testable hypotheses concerning generic drug substitution from existing theoretical models. Many of these models

posit patients' health benefit as an argument in the physician's objective function. Thus, a physician (i) who prescribes a drug (d) to a patient (j) at time (t) has utility

$$\begin{aligned} V_{ijdt} &= \alpha_i \left[\pi_{idt} - e_{ijdt} \right] + \beta_i \left[b_{jd} - \theta_{jdt} p_{dt} u'\{Y_{jt}\} \right] - \gamma_i \left[(1 - \theta_{jdt}) p_{dt} \right] \quad (4.1) \\ \text{with} \quad & \pi_{idt} = f_{dt} + v_{dt} p_{dt} + \eta_{idt}. \end{aligned}$$

Here, π_{idt} denotes the contribution to physician income. As explained in Section 4.3.2, it consists of a fixed lump sum (f_{dt}), a price-dependent component ($v_{dt}p_{dt}$), and an unobserved discount that is the outcome of an individual bargaining process between the physician and the pharmaceutical company (η_{idt}). For the reasons listed in Section 4.3.2, we assume that both discounts and total contributions to physician incomes are higher for generic than for brand-name drugs.

The effort (in money terms) associated with prescribing is denoted e_{ijdt}. In keeping with the literature cited in Section 4.2, this effort is higher for a generic ($d = g$) than a brand-name ($d = b$) drug. For simplicity, the cost of prescribing b is normalized to zero ($e_{ijbt} = 0$). The higher prescribing effort for generic drugs stems from two main sources. First, the physician needs to gather personal experience with the generic drug, which she has already collected for the brand-name drug during the period of patent protection. This cost decreases over time, hence the dependence on time index t. Still, every patient is different, making matching patients with drugs challenging even after an initial information effort. Second, the physician needs to convince the patient that the lower-priced generic drug is not of lower quality. Otherwise, the patient might suffer a 'putative loss' in the sense of Griliches and Cockburn (1994), which might jeopardize the physician's reputation. This cost also declines over time as patients become acquainted with the generic drug. The parameter $\alpha_i > 0$ in Equation (4.1) denotes the weight the physician attaches to the drug's contribution to income. It may well differ between GPs and specialists.

The second term of Equation (4.1) symbolizes net patient benefit. Therefore, a weight $\beta_i > 0$ (with no systematic difference between GPs and specialists assumed) reflects a consideration for the patient's *total utility* derived from health benefit and

disposable income [Bradley and Lesu (2006), De Jaegher and Jegers (2000)] rather than merely for the patient's *health benefit* [Ellis and McGuire (1986)]. Net patient benefit equals health benefit b_{jd} minus the drug's out-of-pocket price $\theta_{jdt}p_{dt}$, with θ_{jdt} denoting the patient's rate of coinsurance (which can be drug-specific) and p_{dt}, the price of the drug. The patient's utility from consuming other goods is $u\{Y_{jt}\}$, which is increasing and concave in patient's income Y_{jt} as well as additively separable from health. Since copayment for a single drug $\theta_{jdt}p_{dt}$ is small in our context, multiplying it by $u'\{Y_{jt}\}$ yields a good approximation of its impact on patient utility. As low-income patients have a high marginal utility of income, they suffer a particularly high utility loss from a given drug cost $\theta_{jdt}p_{dt}$. In the remainder of this paper, there will be no difference in health benefits between the brand-name and the generic drugs ($b_{jb} = b_{jg}$) because bioequivalent drugs are compared (see Section 4.6 for details).

The third term of Equation (4.1) is motivated by agency on behalf of the insurers. Agency towards insurers can be motivated by fear of sanctions or tighter regulation in future.[6] Both types of threats concern GPs and specialists alike. Moreover, high and rapidly increasing health insurance premiums are one of the top concerns of the Swiss population. Therefore, promoting a cost-efficient practice style could create a warm-glow effect of doing what is good for society. Here, $(1-\theta_{jdt})p_{dt}$ symbolizes the cost of the drug treatment falling on the patient's insurer, with $\gamma_i > 0$ indicating the importance of this concern. In view of Equation (4.1), types of (im)perfect agency can be defined as in Table (4.1).

Table 4.1: Types of (Im)Perfect Agency

Types of agency	Parameter values
Perfect agency	$\alpha_i = 0, \quad \beta_i > 0, \quad \gamma_i > 0$
Imperfect agency on behalf of patients	$\alpha_i > 0, \quad \beta_i > 0, \quad \gamma_i \geq 0$
Imperfect agency on behalf of insurers	$\alpha_i > 0, \quad \beta_i \geq 0, \quad \gamma_i > 0$
Lack of agency	$\alpha_i > 0, \quad \beta_i = 0, \quad \gamma_i = 0$

[6]The Swiss health insurers' association (Santesuisse) scrutinizes physicians who exhibit inexplicably high cost of treatment compared to their peers and occasionally sues them.

The generic drug is prescribed if $V_{ijgt} > V_{ijbt}$, hence

$$V_{ijgt} - V_{ijbt} = \alpha_i \left[\pi_{igt} - \pi_{ibt} - e_{ijgt} \right] + \beta_i \left[(\theta_{jbt} p_{bt} - \theta_{jgt} p_{gt}) u'\{Y_{jt}\} \right] + \gamma_i \left[(1 - \theta_{jbt}) p_{bt} - (1 - \theta_{jgt}) p_{gt}) \right] > 0. \quad (4.2)$$

Physician agency can now be analyzed with the help of Equation (4.2). To begin with, non-dispensing physicians do not obtain income from drug prescription ($\pi_{igt} = \pi_{ibt} = 0$), while dispensing physicians are likely to receive a higher income contribution from generic than from brand-name drugs ($\pi_{igt} > \pi_{ibt} > 0$, see Section 4.3.2).[7] PD is therefore expected to increase the prescription of generic drugs.

Hypothesis 1: Given imperfect or lack of agency, dispensing physicians are more likely to prescribe a generic drug compared to non-dispensing ones due to its higher income contribution.

Recall that due to bioequivalence, drug choice affects patient utility exclusively through differences in coinsurance. According to Equation (4.2), both perfect and imperfect patient-related agency thus leads to the prediction that generic drugs are prescribed more often to patients with a high rate of coinsurance (high θ_{jdt}) or low income (high marginal utility of income, $u'\{Y_{jt}\}$), than to other patients.

Hypothesis 2: Given imperfect agency on behalf of patients, generic drugs are prescribed more often to patients with higher rate of coinsurance as long as the brand-name drug is more expensive than the generic, $p_{bt} > p_{gt}$.

Hypothesis 3: Given imperfect agency on behalf of patients, generic drugs are prescribed more to patients with lower incomes because of their higher marginal utility of income.

For the decision whether or not to prescribe a generic drug, only the sign of Equation (4.2) is relevant. If the first term of Equation (4.2) is zero (as for all non-dispensing

[7] In fact, non-dispensing physicians get a fee (TARMED) for prescribing a drug, which however does not differ between brand-name and generic drugs. This fee is therefore irrelevant to our analysis.

physicians), the second term becomes relatively more important for the determination of its sign. Therefore, to the extent that agency motivates physicians to prescribe generic drugs, the effect of patient coinsurance should be more marked for non-PD providers.

Hypothesis 4: Given imperfect agency on behalf of patients, patients' rate of coinsurance is more influential if the physician does not dispense drugs on his or her own account.

Many models of physician agency neglect the third term of Equation (4.2). However, if the influence of copayment represented by $[(\theta_{jbt}p_{bt} - \theta_{jgt}p_{gt})u'\{Y_{jt}\}]$ is low and $(\pi_{igt} - \pi_{ibt})$ is zero, as applies to non-PD providers, all that remains is the (extra) effort of prescribing the generic e_{ijgt}. Therefore, non-PD providers who treat patients with low coinsurance or high incomes should have a very low propensity to prescribe generics due to their higher cost of effort. It takes agency towards the payers of health care [$\gamma_i > 0$ in Equation (4.2)] to make them prescribe a generic.

Hypothesis 5: Given agency on behalf of insurers, non-PD providers prescribe generic drugs to some degree.

In addition to the standard fee-for-service arrangement, Swiss insurers may also offer policies with managed care-type restrictions. Most of these arrangements are aimed at increasing the cost-consciousness of physicians, either by introducing provider cost sharing or by selectively contracting physicians based on indication of efficiency. In both cases, these arrangements are expected to align the interests of physicians with those of the insurers, resulting in an increased influence of the price difference $(p_{bt} - p_{gt})$ on physicians in managed care-settings.

Hypothesis 6: Physicians working in managed care-type settings prescribe more generic drugs because of their increased consideration of the cost of care.

A limitation of our model is that it focuses on physician utility only. This is justified to the extent that asymmetric information about treatment options makes patients

delegate their decision-making authority to physicians. However, this delegation is unlikely to be complete in practice. If patients play a more active role, observed choices are the outcome of a bargaining process between them and physicians [Ellis and McGuire (1990)]. It is important to keep this limitation in mind when interpreting the empirical results in Section 4.7. For example, the patient's rate of coinsurance may impact drug choice not only because of physician agency (as our model suggests), but also because of the patients' own actions.

4.5 Econometric Specification

We estimate the choice between brand-name and generic drugs using a binary choice model. The dependent variable takes on the value one if the physician prescribes g and zero otherwise. Following Ben-Akiva and Lerman (1985), the physician's utility is split into a deterministic and a random component, i.e. $U_{ijdt} = V_{ijdt} + \varepsilon_{ijdt}$, where ε_{ijdt} is unobserved by the researcher. A physician prescribes drug g instead of b if and only if $U_{ijgt} > U_{ijbt}$. Hence, the probability of physician i prescribing g to patient j at time t is given by

$$P_{ijgt} = Pr(V_{ijgt} + \varepsilon_{ijgt} > V_{ijbt} + \varepsilon_{ijbt}) = Pr(V_{ijgt} - V_{ijbt} > \varepsilon_{ijbt} - \varepsilon_{ijgt}) \tag{4.3}$$

with $V_{ijgt} - V_{ijbt}$ given by Equation (4.2). If we assume the random term $\varepsilon_{ijt} \equiv \varepsilon_{ijbt} - \varepsilon_{ijgt}$ to have a logistic distribution, we get the logit choice probability

$$P_{ijgt} = \left(1 + e^{-(V_{ijgt} - V_{ijbt})}\right)^{-1} \tag{4.4}$$

which permits to derive and interpret odds ratios. The drawback of the logit model compared to the probit is that no simple estimators are available as soon as a physician-specific random effect is included. In the probit model, the linear combination of the normal error term and the normal random effect results in a normal distribution. This is not the case for the logit model [see Wooldridge (2002), Ch.15]. By including a physician-specific error term, we allow for within correlation among

the observations belonging to the same physician while still assuming independence of observations across physicians. The physician-specific error captures unobserved factors that we are not able to control for [see also Lundin (2000)]. Examples of unobserved factors that may affect drug choice are favorable experience with a specific drug or the impact of pharmaceutical sales representatives visiting the physician. Therefore, we extend the random utility model above to allow for a physician-specific random effect, i.e. $U_{ijdt} = V_{ijdt} + \nu_i + \varepsilon_{ijdt}$. If $\nu \sim N(0, \sigma_\nu^2)$ one obtains the one-level random-effects logit model [see Wooldridge (2002), Ch.15], with the share of total variance contributed by physician-level variance given by $\rho = \sigma_\nu^2/(\sigma_\nu^2 + \sigma_\varepsilon^2)$ where σ_ε^2 denotes the variance of the overall error term. In addition, one could allow for patient-specific random effects by nesting them with physician-level ones, resulting in a two-level hierarchical regression model [also called mixed-effects model, see Rabe-Hesketh et al. (2001)]. While theoretically attractive, the mixed-effects model could not be estimated due to the complexity of the estimation equation and the size of the dataset.[8] Therefore, we estimated the one-level random-effects model discussed previously. Testing the importance of the physician-specific error term using a likelihood ratio test showed that the one-level random-effects model performed better than the pooled logit regression.

To estimate the coefficients of interest, the systematic component of the utility function $(V_{ijgt} - V_{ijbt})$ needs to be specified. Unfortunately, it is not possible to unambiguously relate the variables of the theoretical model to observed quantities. Still, it is possible to test all the hypotheses that were stated in Section 4.4. The assignments are displayed in Table (4.2).

As explained in Section 4.3.2, we cannot observe the *true* income contribution from physician dispensing, but we expect it to be higher for generic than for brand-name drugs [$\pi_{igt} - \pi_{ibt} > 0$ in Equation (4.2)]. Therefore, we can only include a dummy that indicates whether or not a physician earns an income contribution from dispensing

[8] The mixed-effects model did not converge using Stata 10.

Table 4.2: Overview of the Variables Used for Hypothesis Testing

Variable	Term No. in Equation (4.2)	Hyp. No.	Exp. sign	Confirmed?*
Physician dispensing (PD)	1	1	+	Y (O,A)
General Practitioner (GP)	1	n.a.	+	Y
Interaction of PD and GP	1	1	+	Y
Deductible category (DED2, DED3)	2	2	+	Y (O,A)
Interaction of PD and DED2, DED3	2	4	-	N
Increased rate of coinsurance (COINS)	2	2	+	Y
Interaction of PD and COINS	2	4	-	N
Extra hospital insurance (HOSP)	2	3	-	Y
Accident coverage (ACC)	2	3	-	Y (O,A)
High income area (HIA)	2	3	-	Y (O,A)
Price difference (P)	3	5	+	N (Y for O)
Interaction of PD and P	2,3	n.a.	-	N (Y for O)
HMO contract (HMO)	3	6	+	Y
Gatekeeping contract (GATE)	3	6	+	Y
Control variables: six area types, 25 cantonal dummies, complementary insurance, time trend, patient age and sex, dosage, prescriptions per patient, year of first prescription				

*See Section 4.7.

($PD_{it} = 1$). We expect the coefficient pertaining to the income contribution to be positive, implying that PD increases the probability of choosing g.

The information cost (e_{ijgt}) in Equation (4.2) cannot be measured and thus is absorbed by the random term. A dummy for general practitioners (GP) is interacted with PD to test for systematic differences in α_i of Equation (4.2), i.e. whether GPs react in a different way to the financial incentives from PD than specialists do. A positive interaction effect is expected due to the lower average income of GPs and hence higher marginal utility of income.

Copayment borne by patients is known from the patient's health insurance policy on the one hand and the drug-specific rate of coinsurance on the other. As explained in Section 4.3.3, policies differ in terms of deductibles (DED). Physicians acting as agents [$\beta_i > 0$ in Equation (4.2)] would want to keep patients' out-of-pocket cost low. The higher DED, the more they are expected to prescribe the cheaper generic

(Hypothesis 2). In formulating this hypothesis, DED is viewed as exogenous. Admittedly, high deductibles are typically chosen by higher-income individuals, making θ_{jdt} a function of $u\{Y_{jt}\}$ in Equation (4.2). However, the dataset lacks information that would permit to control for this relationship. Hypothesis (2) can be detailed further. Before January 2006, drug expenditure in excess of DED was subject to a 10 percent coinsurance rate regardless of type g or b. A natural experiment is provided by the policy change of 2006, when the coinsurance rate for (some) brand-name drugs was increased from 10 to 20 percent while it stayed at 10 percent for generics. Producers of brand-name drugs can escape the increased rate of coinsurance by lowering their prices, which is observed in our dataset (see Section 4.6). The effect of the patient's rate of coinsurance on drug choice can be tested by including a dummy COINS that is one if the prescribed drug faces the increased rate of coinsurance at the time of purchase and zero otherwise. In addition, an interaction term PD·COINS serves to test for the influence of financial incentives on physician agency. According to Hypothesis 4, its coefficient is predicted to be negative, indicating less additional generic substitution in the case of physician dispensing.

The hypothesis that generic drugs are prescribed less to patients with higher income due to their lower marginal utility of income (Hypothesis 3) is tested by including dummies for residence in a high-income area (HIA), the purchase of extra hospital insurance (HOSP), and the purchase of accident insurance (ACC). Accident coverage is inversely related to labor force participation because it is usually provided by the employer rather than the health insurer. It thus may be interpreted as an indicator of high income, causing less prescription of generics according to Hypothesis 3.

As to the third term of Equation (4.2), Hypothesis 5 (bearing on γ_i, the role of agency on behalf of insurers) can be tested using the price difference between the brand-name and generic drug ($p_t = p_{bt} - p_{gt}$), to be detailed below. Concerning the relevance of this agency, the following argument can be made. Beyond the deductible, the price difference borne by patients is very small compared to average income. Thus, it is unlikely that consideration for the patients' coinsurance [second term in Equation

(4.2)] provides enough motivation for most of non-dispensing physicians to bear the greater cost of prescribing generic drugs (e_{ijgt}). Therefore, the fact that the market share of generic drugs in our dataset is substantial in the non-PD setting (see Table 4.3) supports the view that $\gamma_i > 0$ in Equation (4.2), suggesting that physicians do consider the cost to insurers when choosing a drug. The interaction term PD·P is used to test whether physician agency is weakened by physician dispensing. As the price difference is part of both the second and the third term of Equation (4.2), both agency on behalf of insurers or agency on behalf of patients could be affected here.

For calculating the price difference, note that it has to be calculated for each combination of package size and dosage, with p_{gt} denoting the average price of N generic products each time. Further, since prices are subject to change, the price difference for a specific size-dosage combination has to be calculated for each month t, i.e. $p_t = p_{bt} - (\sum_n p_{nt})/N \; \forall n = g$. For some of these combinations, only one version is available and no price difference can be calculated. These observations are excluded from the regression analysis. This is not a problem because a prescriber who needs this specific amount of pills and dose does not have a choice between b and g.

For testing Hypothesis 6, differences in health insurance policies can be exploited. Apart from conventional fee-for-service contracts with varying deductibles, consumers can opt for a Health Maintenance Organization (HMO) or a gatekeeping alternative (GATE). In the HMO setting, physicians are paid by capitation rather than the usual fee-for-service. The gatekeeping arrangement uses fee-for-service payments but requires patients to obtain a referral from their general practitioner (chosen from a list issued by their insurer) before seeing a specialist. Moreover, patients in a gatekeeping plan are required to ask for generic drugs. Hypothesis 6 states that both kinds of arrangements should lead to increased consideration of the cost of care by prescribing physicians [higher γ_i in Equation (4.2)] and hence more generic drugs being prescribed. However, it is important to note that patients choosing these contracts are likely less risk-averse and more price sensitive than patients opting for the standard fee-for-service setting. These differences relate to the second rather

than third term of Equation (4.2) yet also contribute to more generic drugs being prescribed.

We complete the econometric specification by a few control variables. Because we expect a positive time trend in favor of generic drugs as practitioners get more familiar with them, we include a variable for the time trend. Patient age and gender serve to control for demographic effects. Also, political attitudes and institutions vary between cantons. In some, PD is widely accepted or even desired while in others, it is disputed. Moreover, unobserved detailing effort by pharmaceutical companies likely differs between cantons. This calls for the inclusion of 25 cantonal dummies, with Zurich constituting the reference category. Individuals can also purchase complementary insurance that covers additional procedures (such as traditional Chinese medicine or otherwise uncovered drugs). These dimensions of complementary insurance likely reflect risk aversion on the part of consumers, making them eschew drug substitution because they are less familiar with the generic alternative.

Drug substitution may also depend on dosage and package size. The reason is that the unobserved contribution to physician income could vary with these two parameters. Therefore, total prescribed dose (number of pills times dosage per pill) is included in the regression. The number of prescriptions per patient controls for long-run chronic patients. Because there is a high likelihood that a patient initiated with a given variety of the drug remains with it, two dummies indicate whether the patient's first prescription took place in 2006 or 2007, when the higher coinsurance rate was already in place.

The deterministic part of the utility function is estimated as

$$\begin{aligned} V_{ijgt} - V_{ijbt} \;=\; & b_0 + b_1\mathrm{PD} + b_2\mathrm{GP} + b_3\mathrm{PD}\cdot\mathrm{GP} + b_4\mathrm{DED2} + b_5\mathrm{PD}\cdot\mathrm{DED2} \\ & + b_6\mathrm{DED3} + b_7\mathrm{PD}\cdot\mathrm{DED3} + b_8\mathrm{COINS} + b_9\mathrm{PD}\cdot\mathrm{COINS} \\ & + b_{10}\mathrm{HOSP} + b_{11}\mathrm{ACC} + b_{12}\mathrm{HIA} + b_{13}\mathrm{P} + b_{14}\mathrm{PD}\cdot\mathrm{P} \\ & + b_{15}\mathrm{HMO} + b_{16}\mathrm{GATE} + b_x X, \end{aligned} \tag{4.5}$$

where the b's are the parameters of interest, X denotes the vector of control variables, and b_x the vector of coefficients of the control variables.

4.6 Data

4.6.1 Chemical Agents Selected

The data was provided by a major Swiss health insurer covering about 15 percent of the Swiss population. They relate to the years 2005 to 2007. The chemical agents selected for analysis are omeprazole (O), amlodipine (A), and ciprofloxacin (C).[9] Omeprazole is used to treat gastric and duodenal abscesses; amlodipine is a calcium channel blocker for the treatment of angina; ciprofloxacin is used to treat specific bacterial infections. Their choice can be justified on the grounds that they have many bioequivalent generic competitors that are available on the Swiss market.[10] Furthermore, these agents belong to the therapeutic categories with substantial sales volume, causing the number of prescriptions in the data to be high. We observe 183,874 (O), 143,358 (A), and 95,580 (C) prescriptions where exactly one package was sold.

The shares of the three brand-name drugs in the sample are depicted in Figure (4.1) for 33 months, starting from March 2005. They dropped throughout 2005, quite likely because prescribing physicians anticipated the increase of coinsurance for certain brand-name drugs from 10 to 20 percent effective January 2006. The new rate was to apply to brand-name drugs whose sales price was 20 percent higher than the cheapest therapeutically equivalent generic.[11] During the first months of 2006, this was the case for all three agents. However, the brand-name producers of amlodipine and ciprofloxacin lowered their prices in month 20 (August, 2006) in order to avoid the extra copayment. In month 29 (May 2007), the producer of the brand-name for

[9] ATC-code: omeprazole (A02BC01), amlodipine (C08CA01), ciprofloxacin (J01MA02). For more details about the investigated agents see *www.drugbank.ca/drugs*.

[10] Number of generics available on the Swiss market (2005–2007): omeprazole (11), amlodipine (12), ciprofloxacin (11).

[11] This is regulated by national law (specifically paragraph Art.38a KLV).

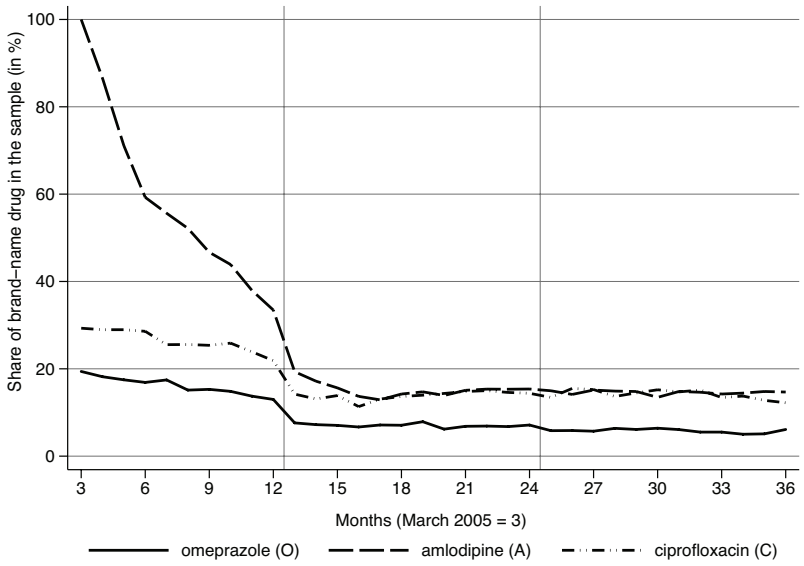

Figure 4.1: Share of brand-name drug between March, 2005 and December, 2007

omeprazole lowered its prices as well, but only for the most commonly prescribed dose (10 mg).

As to amlodipine, the brand-name drug (Norvasc®) went off patent in the spring of 2005, causing it to lose its monopoly position. Since then, the generic Amlodipin-Mepha® has expanded its share in the sample from 18 to 37 percent (2006) and to 38 percent (2007), respectively.

Table 4.3: Sample shares and sales volumes of generic and brand-name drugs, March 2005 - December 2007

	Omeprazole		Amlodipine		Ciprofloxacin	
	PD	Non-PD	PD	Non-PD	PD	Non-PD
Sample share of generics	94%	89%	82%	66%	86%	79%
Sales of generics (in CHF, mn.)	6.3	9.2	3.7	3.5	2.0	1.7
Sales of brand-names (in CHF, mn.)	1.0	2.8	1.5	3.1	0.4	0.6

4.6.2 Physician and Patient Descriptors

In the data set, there are 7,441 physicians prescribing O, 5,995 prescribing A, and 7,693 prescribing C, respectively (the three subsets are overlapping); the share of PD

varies between 43 and 54 percent from March 2005 to December 2007. With 78 to 88 percent, the majority of the prescribers are GPs rather than specialists.

Table 4.4: Descriptive statistics, mean (MN), median (MD), and standard deviation (SD)

	Omeprazole			Amlodipine			Ciprofloxacin		
	MN	MD	SD	MN	MD	SD	MN	MD	SD
Physician dispensing	0.43	0.00	0.50	0.47	0.00	0.50	0.54	1.00	0.50
General practitioner	0.84	1.00	0.37	0.88	1.00	0.32	0.78	1.00	0.42
Patient's deductible	406	300	297	386	300	246	477	300	413
Increased rate of coins.	0.72	1.00	0.45	0.21	0.00	0.41	0.20	0.00	0.40
Extra hospital insurance	0.22	0.00	0.42	0.27	0.00	0.44	0.25	0.00	0.43
Accident insurance	0.75	1.00	0.43	0.83	1.00	0.37	0.68	1.00	0.47
High-income area	0.03	0.00	0.16	0.03	0.00	0.16	0.03	0.00	0.17
Urban area	0.38	0.00	0.49	0.36	0.00	0.48	0.40	0.00	0.49
Suburban area	0.26	0.00	0.44	0.27	0.00	0.44	0.25	0.00	0.43
Average price difference	102	71	75	28	11	30	12	8	9
HMO contract	0.03	0.00	0.16	0.02	0.00	0.13	0.02	0.00	0.14
Gatekeeping contract	0.05	0.00	0.22	0.05	0.00	0.22	0.06	0.00	0.24
Complementary insurance	0.87	1.00	0.33	0.89	1.00	0.32	0.90	1.00	0.31
Patient's age (in years)	62	64	17	70	72	12	58	61	19
Patient's sex (male=1)	0.38	0.00	0.49	0.48	0.00	0.50	0.40	0.00	0.49
Total dosage (in 100 mg)	9.99	11.20	5.90	6.20	5.00	2.80	61.26	50.00	28.60
Prescriptions per patient	7.84	6.00	7.55	8.05	8.00	4.02	2.83	2.00	3.78
First prescription in 2006	0.35	0.00	0.48	0.36	0.00	0.48	0.36	0.00	0.48
First prescription in 2007	0.39	0.00	0.49	0.38	0.00	0.48	0.36	0.00	0.48
Share of prescriptions in 2006	0.35	0.00	0.48	0.36	0.00	0.48	0.36	0.00	0.48
Share of prescriptions in 2007	0.38	0.00	0.49	0.38	0.00	0.49	0.36	0.00	0.48

Note: The prescription is the unit of observation.

The median deductible is the lowest possible (CHF 300). During the study period, about 70 percent of omeprazol prescriptions were subject to the increased rate of coinsurance while this share was lower for amlodipine and ciprofloxacin with shares of 21 and 20 percent, respectively. The share of consumers with extra hospital coverage lies between 22 and 27 percent. The majority of physicians have their practice in urban (36-40 percent) or suburban (25-27 percent) areas while only 3 percent are located in high-income areas. The average savings per prescription for a patient or insurer due to the substitution of the brand-name by a generic counterpart is highest for O with CHF 102, followed by CHF 28 and CHF 12 for A and C, respectively.

The share of insured with an HMO policy varies between 2 and 3 percent, of those with a gatekeeping policy, between 5 and 6 percent. In contrast, between 87 and 90 percent of the insured had signed up for at least one voluntary extra option to broaden the scope of reimbursed services. High shares of 68 and 83 percent have purchased accident insurance. Both the 61,825 patients receiving O and the 27,080 patients receiving C have an average age of about 60 years, and 40 percent are male. The 58,489 patients obtaining A have an average age of 70 years, and 48 percent are male. Ciprofloxacin is prescribed with an average total dosage per prescription of 6,126 mg, compared to a dosage of 999 mg for O and 620 mg for A. On average, a patient receives 8 prescriptions if in need of O or A. In contrast, C is prescribed three times per patient on average. Observations are distributed equally over the three years, with about one third of prescriptions taking place per year. Also, the number of patients starting medication is roughly constant over the years.

4.7 Estimation Results

The odds ratios[12] (ORs) and standard errors resulting from the random-effects logit model described in Section 4.5 are displayed in Table (4.5). The physician-specific variance component contributes 50 to 70 percent of the total error variance, and a likelihood-ratio test clearly speaks in favor of the random-effects specification. The physician-specific variance component is higher than the 40 percent reported by Lundin (2000) and 29 percent reported by Hellerstein (1998). A possible explanation is that some physicians in our dataset only have a small number of patients, the data coming from one insurer only. Moreover, the available information does not permit to distinguish between part-time and full-time, female and male, and younger and older physicians. Coscelli (1998) also mentions considerable physician-specific components in unexplained variance.

[12]The concept of odds ratios and their calculation in the presence of interaction terms can be found in Hosmer and Lemeshow (2000).

4.7.1 Testing for the Influence of Physician Dispensing

Hypothesis 1 predicts that physician dispensing (PD) increases the likelihood of generic prescription. It is tested by Model 1, with physician and patient characteristics controlled for. Additional hypothesis testing calls for interaction terms involving PD and patient characteristics which are added in Model 2 (to be discussed in Section 4.7.2). Therefore, the coefficient of PD in Model 1 shows the average OR across physician and patient groups. In the case of O, it amounts to 3.0 (2.6, 3.4), with the parentheses indicating its 95% confidence interval.[13] For a detailed discussion of its calculation, see Norton et al. (2004) and Garrett (1997). The OR indicates that if the drug is sold on the physician's own account, the odds of generic substitution are three times higher no matter whether the prescriber is a GP or a specialist. For all three agents, the likelihood of generic substitution is around twice as high among GPs than among specialists. Moreover, the interaction between PD and GP yields a positive and significant coefficient in the case of A and C. This could be a sign that GPs with their lower average income, hence higher marginal utility of income, are more influenced by the income contribution of PD than their specialized colleagues. In the case of O, the interaction of PD and GP was insignificant and therefore excluded from the estimation.

The effect of (PD·GP) cannot be inferred from the interaction coefficient directly but needs to be calculated according to the different categories [see Norton et al. (2004)]. In present case, it is given by $exp(\hat{\beta}_{PD})$ for specialists and $exp(\hat{\beta}_{PD}+\hat{\beta}_{PD\cdot GP})$ for GPs.[14] For amlodipine, PD has an OR of 2.4 (1.9, 2.9) for specialists and 3.7 (3.4, 4.1) for GPs, indicating that physician dispensing has a much stronger effect among GPs than among specialists. In the case of C, the discrepancy between GPs and specialists is even stronger. Dispensing specialists reveal a negative PD effect

[13]The 95% confidence interval is calculated according to $CI = exp(\hat{\beta} \pm 1.96 \cdot \widehat{SE}(\hat{\beta}))$, where $\hat{\beta}$ is the logit coefficient. Because Table (4.5) shows ORs, the reader can calculate the necessary quantities according to $\hat{\beta} = ln(\widehat{OR})$ and $\widehat{SE}(\hat{\beta}) = \widehat{SE}(\widehat{OR})/\widehat{OR}$ using the values from the table.

[14]The standard error of this expression is $\widehat{SE}(\hat{\beta}_{PD} + \hat{\beta}_{PD\cdot GP}) = \sqrt{\widehat{Var}(\hat{\beta}_{PD}) + \widehat{Var}(\hat{\beta}_{PD\cdot GP}) + 2\widehat{Cov}(\hat{\beta}_{PD}, \hat{\beta}_{PD\cdot GP})}$, and the confidence intervals for the odds ratios are calculated by $CI = exp(\hat{\beta}_{PD} + \hat{\beta}_{PD\cdot GP} \pm 1.96 \cdot \widehat{SE}(\hat{\beta}_{PD} + \hat{\beta}_{PD\cdot GP}))$.

with an OR of 0.7 (0.6, 0.8), while GPs again exhibit a positive PD effect on generic substitution with an OR of 2.9 (2.6, 3.3). All the OR values discussed have confidence intervals that do not include 1 and thus are significant.

In summary, Hypothesis 1 receives a good deal of support, permitting one to conclude that physician dispensing increases the likelihood of generic substitution due to its higher contribution to physician income. This conclusion holds regardless of whether prescribers are GPs or not and for all of the three chemical substances analyzed, with the one exception of specialized physicians prescribing C. However, it should be noted that there may be additional reasons for dispensing physicians to choose the cheaper generic drug. First, storage entails capital user cost, which is lower for cheap generics. Second, dispensing physicians may be better informed about availability and prices of generics than non-dispensing physicians because of especially targeted marketing activities. Unfortunately, these effects cannot be analyzed with the available data. Still, PD is associated with increased generic substitution. It contributes to lower pharmaceutical expenditure as long as it does not go along with an increase in drug use through supplier-induced demand. This qualification is not addressed here but is analyzed in other recent work. In particular, Rischatsch (2011) analyzes whether dispensing physicians optimize their income contribution from drug dispensing by selling smaller packages, while Chapter 5 of this thesis discusses the impact of physician dispensing on total expenditure for drugs, general practitioners' services, specialists' services and hospital services.

4.7.2 The Role of Physician Agency on Behalf of Patients

To the extent that physicians take the consequences of their prescriptions for the utility of their patients into account, Hypothesis 2 predicts a positive relationship between copayment and generic substitution. Patients with a higher deductible face a higher expected level of copayment; therefore, they should be more likely to receive the generic alternative. The empirical evidence comes from the coefficients of DED2 and DED3 in Model 2 of Table (4.5). In the case of O, the ORs for DED2

and DED3 indicate that a higher deductible increases the likelihood of generic substitution. Patients with a deductible in excess of the legal minimum are two times as likely to receive a generic drug, which supports Hypothesis 2. A stronger effect for DED3 compared to DED2 could not be found for *O*, however. For *A*, the ORs increase from the lowest to the highest deductible category, but only the OR for DED2 is statistically significant. The tendency is the same for *C* but the effect is insignificant. The dummy variable indicating the 2006 increase in coinsurance for expensive brand-names (COINS) is strongly positive for all chemical agents, again supporting Hypothesis 2 (see Table 4.2).

Hypothesis 3 revolves around patient income, stating that richer patients are less likely to receive the generic drug. In Table (4.5), three indicators are used, viz. the purchase of extra hospital insurance, accident insurance, and residence in a high-income area. As to the first indicator, the OR values are consistently below one, indicating that generic drug substitution indeed is less likely. The same is also true for patients with accident insurance and from high-income areas in two of the three cases (*C* is the exception with a negative but insignificant effect). Therefore, there is some supporting evidence for Hypothesis 3 (see Table 4.2 again).

Hypothesis 4 predicts that patients' rate of coinsurance is less influential in the PD mode than in the pharmacy mode. To test it, Model 2 contains interactions between the DED dummies and PD. The interaction terms are generally negative, but only the medium category for *O* is significant, giving some support to Hypothesis 4. Here, the OR for DED2 is 2.2 (1.8, 2.7) for non-PD and 1.6 (1.1, 2.1) for PD. Evidence contradicting Hypothesis 4 comes from *A*, where the interaction effect PD·DED2 is positive and significant but the main effect DED2 is insignificant, leading to the conclusion that non-PD providers do not react to a higher deductible but PD providers do. This difference vanishes again at the highest deductible level since PD·DED3 does not reach statistical significance.

A second test of Hypothesis 4 is provided by the interaction of PD with COINS. However, the evidence is inconclusive. For omeprazole, PD·COINS is highly signif-

icant and positive with an OR of 1.9 (1.7, 2.0) among non-PD providers and 2.6 (2.3, 2.8) PD providers, respectively, while for ciprofloxacin, it is weakly significant but negative, suggesting that PD providers react less to the increase in the rate of coinsurance than their non-PD colleagues. No significant difference could be found for amlodipine. Hence, the evidence does not permit to either confirm or reject the notion that drug dispensing weakens physician agency on behalf of the patient.

4.7.3 The Role of Physician Agency on Behalf of Insurers

Hypothesis 5 states that given agency on behalf of insurers, non-PD providers prescribe generic drugs in spite of higher information cost. Therefore, we expect a higher difference between brand-name and generic prices ($p_{bt} - p_{gt}$) to be positively related to the probability of prescribing the cheaper generic drug. While the estimates for O support Hypothesis 5 with a weak positive effect in favor of generics, the estimates for A and C do not because an increase in the price difference lowers the probability of generic substitution slightly. However, there is other evidence hinting at agency on behalf of insurers. In fact, the descriptive statistics in Table (4.3) show that, for the three selected agents, the share of generic drugs is 66-89 percent in our dataset even in the non-PD market. Recall that non-PD providers do not benefit financially from drug choice, while patient coinsurance beyond the deductible is rather limited compared to average income in Switzerland. Therefore, the high share of generic drugs shows that some physicians choose the lower-priced alternative even in situations when neither they nor their patients derive significant financial benefit from it. It takes agency toward the insurers to motivate physicians to prescribe generic drugs despite higher information cost.

Table 4.5: Odds Ratios of Random-effects Logistic Regression of Drug Choice (Dependent Variable: Generics)

	Omeprazole (O)		Amlodipine (A)		Ciprofloxacin (C)	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Physician dispensing (PD)	2.99***	3.34***	2.36***	2.24***	0.71***	0.74***
	(0.18)	(0.26)	(0.25)	(0.24)	(0.06)	(0.07)
General practitioner (GP)	2.12***	2.13***	1.91***	1.92***	2.21***	2.21***
	(0.22)	(0.22)	(0.16)	(0.16)	(0.19)	(0.19)
Interaction of PD and GP			1.58***	1.56***	4.09***	4.08***
			(0.18)	(0.17)	(0.42)	(0.42)
Deductible category DED2[a)]	2.01***	2.22***	1.15**	1.06	1.02	1.07
	(0.17)	(0.23)	(0.07)	(0.08)	(0.05)	(0.07)
Interaction of PD and DED2		0.70*		1.26*		0.90
		(0.13)		(0.17)		(0.08)
Deductible category DED3[a)]	1.95***	2.50***	1.25	1.42**	1.12	1.20
	(0.39)	(0.66)	(0.18)	(0.25)	(0.11)	(0.16)
Interaction of PD and DED3		0.51		0.71		0.85
		(0.21)		(0.20)		(0.17)
Increased coinsurance (COINS)	2.04***	1.89***	4.52***	4.58***	2.14***	2.26***
	(0.08)	(0.08)	(0.10)	(0.13)	(0.09)	(0.11)
Interaction of PD and COINS		1.35***		0.97		0.88*
		(0.07)		(0.05)		(0.06)
Extra hospital insurance (HOSP)	0.68***	0.68***	0.75***	0.75***	0.93**	0.93**
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Accident insurance (ACC)	0.80***	0.80***	0.89***	0.89***	0.97	0.97
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
High-income area (HIA)[b)]	0.47***	0.48***	0.57***	0.57***	0.91	0.91
	(0.11)	(0.11)	(0.09)	(0.09)	(0.17)	(0.17)
Price difference (P, in 10 CHF)	1.03***	1.04***	0.82***	0.81***	0.94**	0.93**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)	(0.03)
Interaction of PD and P		0.97***		1.02***		1.00
		(0.00)		(0.01)		(0.03)
HMO contract (HMO)[c)]	1.94***	1.91***	1.99***	1.99***	1.37***	1.37***
	(0.25)	(0.25)	(0.25)	(0.25)	(0.15)	(0.15)
Gatekeeping contract (GATE)[c)]	2.43***	2.37***	1.63***	1.64***	1.35***	1.35***
	(0.21)	(0.21)	(0.09)	(0.09)	(0.09)	(0.09)
Complementary insurance	1.15***	1.15***	1.17***	1.17***	1.00	1.00
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Time trend (in months)	1.03***	1.03***	1.08***	1.08***	1.05***	1.05***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patient age (in 5 years)	1.01**	1.01**	0.98***	0.98***	0.99***	0.99***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Patient sex (male=1)	1.26***	1.26***	1.14***	1.14***	1.02	1.02
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)
Total dosage (in 100 mg)	0.93***	0.93***	1.09***	1.09***	1.00*	1.00*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Prescriptions per patient	0.94***	0.94***	0.96***	0.96***	0.98***	0.98***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
First prescription in 2006	1.33***	1.32***	1.21***	1.21***	1.23***	1.23***
	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)
First prescription in 2007	1.38***	1.37***	1.04*	1.04*	1.09*	1.09*
	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.05)
Log-likelihood at convergence:	-35,970	-35,918	-51,481	-51,473	-29,390	-29,388
Observations/Physicians:	183,874/7,441		143,358/5,995		95,580/7,693	

Additional regressors: Six area and 25 cantonal dummies. Standard errors in parentheses.

***p<0.01, **p<0.05, *p<0.1. [a)] DED2 = CHF 1,000 or 1,500, DED3 = CHF 2,000 or 2,500.

[b)] basis category is urban area, [c)] reference category is basic insurance.

The interaction PD·P is again used to test whether the financial incentives attached to PD weaken physician agency. The price difference being part of both the second and the third term of Equation (4.2), both agency on behalf of the patient and on the behalf of the insurer can be affected. For O, the price difference has an OR of 1.0 (1.03, 1.05) for non-PD physicians and an OR of 1.01 (1.00, 1.02) for PD physicians, pointing to a weakly negative association of PD and agency. The opposite is observed in the case of A, where the OR pertaining to non-PD providers is 0.81 (0.80, 0.82) and the OR pertaining to PD providers is 0.83 (0.82, 0.83). For C, no significant difference between non-PD and PD providers is observed, with ORs amounting to 0.93 (0.88, 0.99) and 0.94 (0.88, 1.00), respectively. Therefore, the evidence with regard to the interaction of PD and agency is inconclusive.

With respect to Hypothesis 6, the managed-care variables 'HMO' and 'gatekeeping' reveal an increasing likelihood of generic substitution for all three chemical agents, with ORs between 1.4 and 2.0, as predicted (see Table 4.2).

4.7.4 Control Variables

The control variables lead to the following conclusions. In Model 2 of Table (4.5), there is evidence for the expected positive time trend towards generic drugs, a higher likelihood of generics being prescribed to men compared to women, no evidence of the total amount of dosage prescribed having influence on the choice of drug version, and a negative effect of number of prescriptions on the likelihood of generic prescription. Finally, the year when the patient's medication started is important for drug choice and significant for all three chemical agents. Patients who received the first prescription in 2006 are between 1.2 and 1.3 times more likely to be prescribed a generic. In the case of amlodipine and ciprofloxacin, the likelihood for 2007 is higher than for 2005 but lower than for 2006. This could reflect the fact that the two pertinent brand-name producers lowered their price in the interest of a decreased coinsurance rate, enabling them to regain market-share. By way of contrast, the

brand-name producer of omeprazole waited until 2007, causing it to lose market share in both years.

One might criticize that dispensing physicians do not react to an individual patient when choosing between g and b because they have already decided what pharmaceuticals to have in their portfolio. However, they are likely to make this choice anticipating the kind of patients they will face from past visits, causing them to store the drugs that best match their clientele.

4.8 Conclusions

This research analyzes the role of physicians' and patients' financial incentives in the choice between generic and brand-name drugs. Prescribing the generic alternative takes more effort on the part of the physician for two main reasons: First, she needs to acquire information about new drugs which enter the market only after patent expiration of the brand name. Second, she needs to convince the patient that the cheaper generic is not of lower quality. The physician is willing to make this effort only if the benefit from choosing the generic is sufficiently high. Generic drugs have higher benefit because of three reasons, namely financial benefits, agency towards the patient, and agency towards insurers. The influence of these three components is estimated using a large set of drug claims data from Switzerland.

Regarding financial incentives, this data is ideal for analysis because some – but not all – Swiss physicians have the right to dispense drugs on their own account. Physicians with this privilege derive a significant part of their income from the sale of drugs, causing financial incentives associated with drug dispensing to be substantial. Physician dispensing is found to be associated with a higher likelihood of prescribing generic drugs, which is likely due to a higher contribution to physician income in comparison with that of brand-name drugs (Hypothesis 1; see also Table 4.2). A limitation of our analysis is that we are unable to separate this effect from other differences between dispensing and non-dispensing physicians. In particular, information costs for prescribing generic drugs might be lower for dispensing

physicians as they are targeted by sales representatives and may therefore be better informed about availability and prices of drugs than their non-dispensing colleagues. Additionally, dispensing physicians have to finance and manage storage, tying up capital and causing opportunity costs.

Turning to agency towards patients, we test whether physicians respond to the financial burden caused by copayment. Choosing the lower-priced generic drug serves to decrease this burden without affecting the quality of medication due to bioequivalence of the generic substitutes studied here. We find that the likelihood of receiving the generic increases for patients with a higher deductible (Hypothesis 2). In addition, the rate of coinsurance (which applies when the deductible is exceeded) was increased for certain brand-name drugs during our observation period. Although this change caused but a small additional burden per patient compared to income, it does go along with a strongly increased use of generic drugs. A likely contributor is that the government's initiative to promote generic substitution alloyed concerns about quality on the part of both prescribers and patients.

The variation in deductibles and coinsurance permits to study the interaction between physicians' financial incentives and their patient agency. Given imperfect agency on behalf of patients, dispensing physicians are predicted to respond less strongly to a hike in copayment than non-dispensing ones (Hypothesis 4). However, the evidence found in our data is mixed, failing to support the notion that drug dispensing weakens physician agency, as argued by pharmacists' lobbying groups and some Swiss politicians.

Moreover, most of the odds ratios pertaining to proxies of patient income (residence in a high-income area, purchase of extra hospital and accident insurance) suggest that wealthier patients have a higher probability of receiving brand-name drugs because the price difference between them and the generic substitute has less of an effect due to lower marginal utility of income of the wealthy (Hypothesis 3).

Consideration of the savings for insurers might provide an additional motivation for the prescription of the cheaper generic alternative (Hypothesis 5). However, this

effect could be confirmed for only one drug in the econometric estimation (see Table 4.2 again). Nevertheless, the high willingness of non-dispensing physicians to prescribe generic drugs points to some degree of agency towards insurers. Last but not least, physicians working in managed care-type arrangements are found to prescribe more generic drugs than their colleagues, pointing to an increased cost awareness in the managed care setting (Hypothesis 6).

In sum, financial incentives, agency towards the patient, and agency towards insurers are all found to markedly influence generic substitution. Moreover, government initiatives to promote generic drugs can be effective even in the presence of weak financial incentives because they may reassure physicians and patients of the safety and high quality of generic drugs. However, if government were to try to markedly reduce generic prices, it might weaken the incentives for generic substitution, at least for dispensing physicians. The reason is that physicians' financial incentives may encourage rather than undermine generic substitution.

Acknowledgements

The authors would like to express their thanks to Randall Ellis, Albert Ma, Teresa Gibson, Giovanni Mastrobuoni, Harry Telser, and Martin Peter for helpful comments and discussions. Special thanks to the CSS Institute for Empirical Health Economics, especially Konstantin Beck, Ute Studer-Merkle, Silvia Gaehwiler-Burkard, Karin Pohl, Stefan von Rotz, Ute Kunze, Bernhard Keller, and Urs Kaeser for provision of data and many insights. Any errors or omissions are our own.

Prescribers' Responses to Financial Incentives - Theory and Evidence

Maria Trottmann

Abstract: Several countries (Canada and the US amongst others) have recently expanded pharmacists' rights to authorize prescriptions. This amounts to a partial lifting of the traditional separation between the prescription and the sale of drugs. This separation exists in some Swiss jurisdictions but not in others, where physicians have the right to dispense drugs on their own account. Using individual patient records and a standard two-part model specification, this article finds no evidence that dispensing by physicians has a significant effect on either drug expenditure or total health care expenditure. However, dispensing by physicians is associated with an increased use of primary care services and slightly fewer referrals to hospitals.

Keywords: Physician agency, prescription drugs, referral decisions
JEL classification: I10; I18; I19

Chapter 5

Prescribers' Responses to Financial Incentives - Theory and Evidence

5.1 Introduction

Agency relationships in health care occur because of information asymmetries between prescribers, patients and insurers. Patients are generally unaware of the possibilities and consequences of treatment, causing them to delegate medical decisions to prescribers. During the decision-making process, prescribers gain considerable knowledge that cannot be shared with insurers at a reasonable cost ['specific knowledge' in the sense of Jensen and Meckling (1992)]. Therefore, insurers cannot fully verify prescribers' decisions even by hiring medical personnel. This provides prescribers with a considerable level of independence. If prescribers sell medical services with positive markups, their monetary self-interest is frequently in conflict with the other parties' interest for cost-effective care. For this reason, supply conditions of medical care are strictly regulated even in countries claiming to be market economies [Arrow (1963)]. It is however difficult to empirically quantify the extent of this agency problem or the success of the regulation addressing it.

The market for pharmaceuticals is no exception. In most Western nations, drugs cannot be prescribed and dispensed by the same professional because dispensing prescribers may be inclined to prescribe too many, too expensive, or even clinically

inappropriate drugs. However, should a professional be qualified to carry out both tasks, the combination of the two roles has potential advantages for the patient.[1] First, it increases patients' choice of drug provider, thereby increasing competition and responsiveness to patients' preferences. Second, it saves the (time) cost of visiting a second provider. Third, using the same personnel and infrastructure for both tasks may well serve to reduce prices.[2] Fourth, outpatient care providers who earn margins from drug dispensing might have an increased incentive to treat patients instead of referring them to hospitals [Zweifel (1985)]. These advantages must be gauged against the agency problems that have been previously described.

This paper analyzes the consequences of giving physicians the right to dispense drugs on their own account (Physician Dispensing, PD). The data originates from Switzerland where some jurisdictions allow PD while others do not. Physicians who dispense earn a marked proportion of their revenue from this activity, reflecting substantial financial incentives pertaining to prescribing. The innovation in this study is that I am analyzing the impact of PD not only on drug expenditure but also on expenditure for other types of medical care. In a first part, a model of physician decision making serves to formulate three hypotheses. These are then tested using patient-level data of about 240,000 Swiss patients.

The remainder of this paper is structured as follows. Section 5.2 starts off with a short description of the literature. Section 5.3 contains a description of the policy setting. In Section 5.4, a theoretical model of physician decision making is presented and testable hypotheses on the influence of PD on health care expenditure are derived. The estimation strategy is described in Section 5.5, while data and descriptive statistics are provided in Section 5.6. Section 5.7 discusses estimation results while section 5.8 rounds off with a summary and conclusions.

[1]The question of necessary qualifications is not addressed by this paper. It is assumed throughout that there exist drugs that could be dispensed by physicians or prescribed by pharmacists without a reduction of quality or safety.

[2]In a recent report on pharmaceutical regulation, the Swiss government stated the opinion that physicians face lower dispensing cost than pharmacists do [BAG (2009)]. The report concluded that it would be 'unfair' to pharmacists to allow both dispensing channels to compete on prices.

5.2 Literature Review

Several countries recently reformed the division of labor between physicians and pharmacists. In the Anglo-Saxon world, regulators have extended the right of pharmacists to prescribe drugs. Amidst a growing shortage of primary care providers, these reforms are aiming at a better use of pharmacists' skills. In the US for example, at least 25 States grant pharmacists the right to prescribe on the basis of written guidelines issued by the treating physician ('protocol-based prescribing'). Even more States allow pharmacists to issue prescription refills [Emmerton et al. (2005)]. In the United Kingdom, pharmacists have recently been authorized to prescribe drugs to patients who joined a clinical management plan that ensures good communication between the pharmacist and the treating primary care physician [Guillaume and Cooper (2008)]. Furthermore, cooperations between pharmacists and physicians or practice groups are emerging in the US, Canada and Australia [Emmerton et al. (2005)]. The impact of these changes has not yet been analyzed empirically.

In contrast to the Anglo-saxon world, many countries in South East Asia are strengthening the separation between prescribers and sellers of drugs. In traditional Chinese medicine, a formal separation of the two roles was unknown. Indeed, charging fees for personal services was deemed immoral; thus physician remuneration was technically related only to drugs [Ikegami and Cambell (1995)]. In the year 2000 health reform in South Korea, the dispensing of drugs by physicians and drug prescription by pharmacists was outlawed. Descriptive statistics presented by Kim et al. (2004) indicate that the reform was associated with an increase in drug spending, due to the fact that physicians substituted lower-priced, home-grown, generic drugs by more expensive, imported, brand-name drugs. An explanation is that physicians were alienated by the reform and had little incentive to keep health care expenditure low. In Taiwan, Liu et al. (2009) found that clinics on global budgets are more likely to prescribe generics than clinics reimbursed fee-for-service. Iizuka (2007) finds that markups significantly influence the choice of anti-hypertensive drugs by Japanese

physicians. His simulations suggest that reducing prescribers' markups to zero could reduce expenditures in this drug class by as much as 15 percent.

In Switzerland, several studies have analyzed the effects of physician dispensing using aggregate data. Dummermuth (1993) compares two otherwise similar neighboring cantons (Lucerne with PD and Argovia without PD), finding PD to be associated with slightly higher per capita drug expenditure. This finding is in line with Beck et al. (2004), who relate per-capita drug expenditure to PD regulation and other characteristics of cantons. A different result is obtained by Vatter and Ruefli (2003), who control for a very comprehensive set of political and socioeconomic variables. They identify a significantly negative effect of the share of PD providers on per capita HCE. More surprisingly still, Schleiniger et al. (2007) estimate a significantly negative effect of PD on drug expenditure which is robust across several specifications. An early study using physician level data is Zweifel (1985), who finds that physicians who dispense drugs had lower hospital referral rates. However, cost savings pertaining to reduced referrals were not enough to offset the increased cost of drugs. According to a recent study by Busato et al. (2010), PD-physicians bill less services per patient than others, likely due to reduced incentives to generate additional income. Last but not least, it was found in Chapter 4 of this thesis that generic substitution is more common among dispensing physicians than among pharmacists. This is probably due to generic drugs offering higher contributions to income.

5.3 Policy Setting

5.3.1 Swiss Health Insurance

Health insurance in Switzerland is mandatory for a rather comprehensive 'basic' basket of medical services and pharmaceuticals, written by some 80 non-profit insurers competing in a regulated market. Insurers must accept all applicants during semi-annual open enrollment periods. Premiums can be differentiated by area of residence but not by age or health status. Low-income individuals are eligible for premium

subsidies that are funded from general tax revenue. The baseline contract grants insured individuals unlimited access to all licensed physicians and most hospitals in their region of residence. They face a minimum annual deductible of CHF 300 (some EUR 200 as of 2005) and a coinsurance rate of 10 percent up to a cap of CHF 700 (EUR 470) per year. While this system is generally found to ensure access to comprehensive health care to all citizens, it is criticized for high and rapidly increasing HCE, lack of co-ordination between providers, and lack of information on quality and efficiency [OECD (2006)]. In response to these problems, insurers have been granted the right to offer managed-care type options (since 1994) and higher deductibles (since 1996) in return for lower premiums.

5.3.2 Regulation of Physician Dispensing

Physician dispensing (PD) is regulated on the cantonal level. Therefore, a variety of regulations can be observed within Switzerland. In French and Italian speaking regions, PD is unknown except in emergencies and for complicated treatments. In German speaking regions, only one canton has such a rigid regulation (Basel City, BS). Many of the remaining cantons allow physicians to dispense on their own account, while others allow it under specific conditions (for example limited access to pharmacies).

In regions that allow PD, physicians gain on average 18 percent of their revenue from drugs. This amount is higher for general practitioners (28 percent) and lower for specialists (8 percent). Despite these marked financial gains, only 60 - 85 percent of physicians who are entitled to dispense on their own account actually choose to do so [Hunkeler (2008)].[3] Although summary statistics can be misleading, it is interesting to consider Table 5.4 in Appendix 5.9.1. Cantons that rely on PD seem to have lower drug expenditure on average than cantons with pharmacy-based systems. In this study, special attention will be given to Lucerne (LU), Zurich (ZH), Bern (BE), and Argovia (AG). These four cantons are all German speaking and similar in mentality,

[3] In the definition by Hunkeler (2008), a physician dispenses on his own account if his drug turnover exceeds CHF 25,000 per annum.

yet have differing laws regarding PD. Lucerne is the most PD-friendly, granting the right to dispense to all physicians. Zurich (ZH) allows PD except in the two largest cities (Zurich and Winterthur), whilst the canton of Berne (BE) grants permission to dispense to physicians in communities where not more than two pharmacies are located. The canton of Argovia (AG) does not allow PD, except to physicians who practice in underserved municipalities. Lucerne and Argovia are very comparable in terms of socio-economic status, so comparing them in Table 5.4 seems appropriate. Drug expenditure and total health care expenditure are shown to be lower in PD-friendly LU. Physician density is similar in the two cantons, but Lucerne has a greater share in primary care.

5.4 Theoretical Model and Hypotheses

The purpose of this section is to build a model of physician behavior from which testable hypotheses can be derived. The model is tailored to general practitioners (GPs), who account for the lion's share of dispensing physicians in Switzerland [Hunkeler (2008)]. It is assumed that patients who arrive at the GP differ in their need for a referral, denoted (β). The GP knows the distribution $F(\beta)$ ex ante. Before seeing patients, the GP fixes a level ($\hat{\beta}$) up to which she will treat the patient. If the patient's severity β is below her ($\hat{\beta}$), she decides about the amount of drug (d) and non-drug GP care (hereafter: treatment, t) that the patient will consume.

In order to simplify the analysis, no bargaining process between the GP and the patient is modeled. It is instead assumed that the physician takes patient utility into account when she makes her treatment decisions. The model can be split into two decision stages:

1. The GP fixes a level of patient severity ($\hat{\beta}$) up to which she will treat the patient rather than referring him.

2. Patients arrive at the GP. If their β is above $\hat{\beta}$, the GP refers the patient to a hospital. If the β is below $\hat{\beta}$ the GP decides on the quantity of drugs (d) and treatment (t) that the patient will consume.

When taking the referral decision, the GP anticipates the optimal amounts of d and t that she will choose if she decides to treat the patient herself. Therefore, the second stage is discussed first.

5.4.1 Second Stage: Optimal Levels of Drugs and Treatments

If the GP treats the patient herself, she chooses the amount of drugs (d) and treatment (t) that the patient will consume.[4] For each unit d or t, she earns a fee denoted by π^d or π^t respectively. In addition to the monetary profit, the GP derives utility from generating benefit to the patient. The relative weight she attributes to patient benefit compared to monetary profit is denoted by the agency parameter α [see also Ellis and McGuire (1986)]. For simplicity, the two types of care are modeled as perfect substitutes. Patient utility from using drugs is denoted $V^d(d)$ and assumed increasing and concave $\left(\frac{\partial V^d(d)}{\partial d} > 0, \frac{\partial^2 V^d(d)}{\partial d^2} < 0\right)$, and the analogous holds for treatment t $\left(\frac{\partial V^t(t)}{\partial t} > 0, \frac{\partial^2 V^t(t)}{\partial t^2} < 0\right)$.

In keeping with DeJaegher and Jegers (2006), it is assumed that the GP considers the health benefit generated to the patient as well as the financial burden imposed on him by copayments. The patient is partly insured and pays a share (θ) of the cost of care, up to a stop-loss (Θ). To simplify the analysis, the stop-loss is reached if and only if a referral is made.[5] The prices for a unit of drugs or treatments are assumed to be equal and are normalized to 1.

Both services have positive and convex cost to the GP, thus $\frac{\partial C(d,t)}{\partial d} > 0$, $\frac{\partial^2 C(t,d)}{\partial d^2} > 0$, and the analogous for t. Furthermore, resources spent on one type of care cannot be spent on the other, so $\frac{\partial^2 C(t,d)}{\partial d \partial t} > 0$. For later use, it is also assumed that the second

[4] This section draws upon the model presented by Eggleston (2005).

[5] The simplification is justified as only 15 out of over 240,000 patients reached the stop-loss without a referral in the data. Among those with a referral to a hospital, around 60 percent reached the stop-loss.

order cross derivative of the cost function is not larger than the single second order derivative $\frac{\partial^2 C(t,d)}{\partial d \partial t} \leq \frac{\partial^2 C(t,d)}{\partial d^2}$ and $\frac{\partial^2 C(t,d)}{\partial d \partial t} \leq \frac{\partial^2 C(t,d)}{\partial t^2}$. The GP's maximization problem is thus

$$U^{GP}|\text{No Referral} = \pi^t t + \pi^d d + \alpha \left[V^t(t) + V^d(d) - \theta(t+d)\right] - C(t,d). \quad (5.1)$$

FOC:

$$0 = \pi^d + \alpha \left[\frac{\partial V^d(d)}{\partial d} - \theta\right] - \frac{\partial C(t,d)}{\partial d} \quad (5.2)$$

$$0 = \pi^t + \alpha \left[\frac{\partial V^t(t)}{\partial t} - \theta\right] - \frac{\partial C(t,d)}{\partial t} \quad (5.3)$$

Dispensing physicians earn substantially higher revenues per drug quantity prescribed than non-dispensing physicians. Therefore, the right to dispense drugs leads to an exogenous increase in the profits from drugs ($\pi^d \uparrow$) while the profits from physician treatments remain constant (π^t). The impact of this exogenous increase can be calculated by taking the total differential of the first order conditions.[6] Using matrix notation, this reads

$$\begin{bmatrix} \alpha \frac{\partial^2 V^d(d)}{\partial d^2} - \frac{\partial^2 C(t,d)}{\partial d^2} & -\frac{\partial^2 C(t,d)}{\partial d \partial t} \\ -\frac{\partial^2 C(t,d)}{\partial t \partial d} & \alpha \frac{\partial^2 V^t(t)}{\partial t^2} - \frac{\partial^2 C(t,d)}{\partial t^2} \end{bmatrix} \begin{bmatrix} \Delta d \\ \Delta t \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \Delta \pi^d \quad (5.4)$$

The matrix on the left hand side of Equation (5.4) is the Hessian matrix of Equation (5.1). With $|H|$ denoting the determinant of the Hessian, applying Cramer's rule yields

$$\frac{\Delta d}{\Delta \pi_d} = \frac{\frac{\partial^2 C(d,t)}{\partial t^2} - \alpha \frac{\partial^2 V^t(t)}{\partial t^2}}{|H|} = \frac{(+)}{(+)} > 0. \quad (5.5)$$

[6] Amongst others, see Simon and Blume (1994) Ch. 15.3.

For a local maximum, the determinant of the Hessian must be positive and the first leading principal minor, negative. Equation (5.5) leads to the hypothesis that physician dispensing increases drug expenditure.

Hypothesis 1: PD increases drug expenditure.

If the physician agency parameter α increases, the denominator of Equation (5.5) increases faster than the numerator. In consequence, $\frac{\Delta d}{\Delta \pi_d}$ becomes smaller. It is intuitive that high-agency GPs react less to the financial incentives from drug dispensing than low agency GPs.

Due to the fact that time spent on one service cannot be spent on another service, increased drug prescriptions make the provision of physician treatments more costly [see also Eggleston (2005)]. Using Cramer's rule once more, one obtains

$$\frac{\Delta t}{\Delta \pi_d} = -\frac{\frac{\partial^2 C(d,t)}{\partial d \partial t}}{|H|} = -\frac{(+)}{(+)} < 0 \tag{5.6}$$

Hypothesis 2: PD lowers the amount spent on physician treatments.

5.4.2 Referrals

Patients who arrive at the GP differ in their need for a referral (β), which is between [0,1] and drawn independently from distribution $F(\beta)$.[7] If a referral is made, the patient receives a fixed amount $\bar{S}$ of hospital care. The benefit $v^s(\bar{S})$ that is generated to patient health depends linearly on β, i.e. $v^s(\bar{S}) = \beta\bar{S}$. Because the referral was made, the patient's copayment equals the stop-loss of Θ.

The GP is assumed not to earn money for a referral, but she cares about patient benefit. The GP's utility from a referral is thus

$$U^{GP}|\text{Referral} = \alpha[\beta\bar{S} - \Theta]. \tag{5.7}$$

The critical value for a referral ($\hat{\beta}$) is set ex ante, based on expected utility. Whenever an individual β is above $\hat{\beta}$, the GP's utility is given by Equation (5.7). When

[7] See Iversen and Ma (2010) for a similar specification.

β is below $\hat{\beta}$, it is given by the maximization of Equation (5.1). With d^* and t^* denoting the values of drugs and treatment that maximize (5.1), the expected utility maximization problem reads

$$EU^{GP} = \int_0^{\hat{\beta}} U(d^*, t^* | \text{No Referral}) f(\beta) d\beta + \int_{\hat{\beta}}^1 \alpha[\beta \bar{S} - \Theta] f(\beta) d\beta \quad (5.8)$$

$$\text{FOC: } \frac{\partial EU^{GP}}{\partial \hat{\beta}} = U(d^*, t^* | \text{No Referral}) f(\hat{\beta}) - \alpha \hat{\beta} \bar{S} f(\hat{\beta}) + \alpha \Theta f(\hat{\beta}) = 0 \quad (5.9)$$

$$\rightarrow U(d^*, t^* | \text{No Referral}) = \alpha[\hat{\beta} \bar{S} - \Theta] \quad (5.10)$$

Although the utility level in absolute terms is not determined by the model, it can be stated without loss of generality that the extra income from physician dispensing increases physician utility $\left(\pi^d \uparrow, U(d^*, t^* | \text{No Referral}) \uparrow\right)$, hence

$$U^{PD}(d^*, t^* | \text{No Referral}) > U^{NON-PD}(d^*, t^* | \text{No Referral}). \quad (5.11)$$

For Equation (5.10) to hold after an increase in the left-hand side, the physician must increase her $\hat{\beta}$. Therefore, fewer cases are referred to hospitals.

Hypothesis 3: PD lowers the number of referrals.

5.4.3 Physician Income, Demand Inducement, and Agency for the Patient

Two implicit assumptions behind Equation (5.1) deserve attention. First, the physicians' marginal utility of income is assumed to be constant. This assumption implies that the physician's interest in supplying more services is independent from her total income,[8] which contradicts some widely applied models of physician behavior. Indeed, the 'target-income-hypothesis' [Evans (1974)] or its 'income-effect' specification [McGuire and Pauly (1991)] assume that physicians have a strong interest in providing more services (eg. by inducing demand) as long as their income is low. When their income is higher, their interest in providing services becomes weaker. In

[8] Due to the convexity cost function, it is not independent of the total amount of services supplied.

the present setting, PD-physicians earn more income per case compared to non-PD physicians. Therefore, the 'income effect' concept of physician behavior leads to the hypothesis that PD lowers the amount of physician treatments provided. It thereby supports Hypothesis 2.

The second implicit assumption is that the agency parameter α is exogenous and independent from the other parameters in the model. This is a standard assumption in health economic papers, nevertheless, it is often challenged in the political discussion in Switzerland, notably by pharmacists and their lobbying groups, who argue that the extra income from drug dispensing is likely to weaken physician agency for the patients. Although endogenizing the degree of physician agency is beyond the scope of this paper, the argument will be reconsidered in the empirical section.

5.5 Econometric Specification

The hypotheses (1) to (3) from the theoretical model are tested using a large, patient-level claims data set. Two main econometric issues have to be addressed. First, the distribution of health care expenditure is truncated at zero and highly skewed to the right. This is addressed by the application of a two-part model (see Section 5.5.1 and 5.5.2). Second, health care expenditures vary between regions for many reasons, which might be confounded with the effects of PD. This calls for the inclusion of region specific fixed effects (see Section 5.5.3).

5.5.1 Modeling Health Care Expenditure Data

The distribution of health care expenditure data is typically characterized by two facts: First, a considerable part of the population does not use any health care within a given period. Second, the positive spending is strongly skewed to the right with a long tail of few individuals needing very expensive care. An extensive debate in the health economics literature revolves around the estimation approach best suited to this distribution. The most prominent candidates are hurdle models such as the original

Tobit model, sample selection models (SSM) as suggested by Heckman (1979), and the two-part models (2PM) suggested by Duan et al. (1982).

In this application, the two-part model is considered most appropriate. The Tobit model is known to be inconsistent in the presence of heteroskedasticity, which is strong in our data set. The SSM does not require the error terms in the utilization equation to be normally distributed [Olsen (1980)], but it suffers from multicollinearity if the same set of variables is used in the selection and the utilization equation. Several studies use Monte Carlo simulations to show that high multi-collinearity can lead to poorer predictive power of the SSM compared to the 2PM [Manning et al. (1987), Hay et al. (1987), Leung and Yu (1996)].[9] Moreover, Leung and Yu (1996) show that the multi-collinearity problem is aggravated if the censoring is high (ie. if there are 'many' zeros). In appendix 5.9.2, I show test results for multi-collinearity based on the variance inflation factor. In the case of less frequently used types of health care such as hospital care, applying the SSM indeed leads to very severe multi-collinearity problems in our dataset.

The 2PM approach separates the probability of using care from the amount of care if positive. With y representing the outcome variable (health care expenditure) and X representing a matrix of independent variables, the expected value of the outcome becomes

$$E(y|X) = P(y > 0|X) \cdot E(y|X, y > 0). \tag{5.12}$$

The probability $P(y > 0|X)$ is commonly estimated by a probit model. In the second part, Duan et al. (1982) transformed the dependent variable by the log and applied OLS. However, Manning (1998) notes that retransforming the expected value of the logarithmic outcome variable back to raw scale is very challenging in the presence of heteroscedasticity. Alternatively, the second part can be estimated by a generalized linear modeling approach with a log link [Manning and Mullahy (2001)].

[9]The results by Hay et al. (1987) are particulary relevant here because the authors generated their data to mirror Swiss health care expenditure data.

For the specification of the variance function, the two authors proposed a Park test. Applying this test to our data indicated that the error variance is proportional to the squared conditional mean (the gamma structure).

5.5.2 Marginal Effects in the 2PM

In order to assess the marginal effect (ME) of a variable on the expected outcome, the two parts have to be combined. As both equations are nonlinear, the size of the marginal effects depends on the values of all control variables. Let x_1 be a binary variable of interest and X a vector of additional independent variables. In the absence of interaction terms, the marginal effect can be calculated using,

$$\frac{\Delta E[y|X]}{\Delta x_1} = [\Phi(\beta_1 + X\beta)exp(\beta_1 + X\beta) - \Phi(X\beta)exp(X\beta)]. \tag{5.13}$$

If interaction terms are present, calculating the marginal effects is more complicated. For the case of probit or logit models, interaction terms are derived by Ai and Norton (2003) and sketched in the appendix. Following their method, the correspondent formulas for the two-part marginal effects are derived by Frondel and Vance (2009). Let x_1 and x_2 be two interacted dummy variables and X a vector of additional independent variables. Then, the interaction term is [see appendix 5.9.3 for more detail],

$$\begin{aligned}\frac{\Delta^2 E[y|X]}{\Delta x_1 \Delta x_2} = {} & \Phi(\beta_1 + \beta_2 + \beta_{12} + X\beta)exp(\beta_1 + \beta_2 + \beta_{12} + X\beta) \\ & -\Phi(\beta_1 + X\beta)exp(\beta_1 + X\beta) - \Phi(\beta_2 + X\beta)exp(\beta_2 + X\beta) \\ & +\Phi(X\beta)exp(X\beta).\end{aligned} \tag{5.14}$$

The standard errors of expression (5.13) are obtained by bootstrapping.[10] As Ai and Norton (2000) point out, bootstrapping might yield more accurate estimations in

[10] To be specific, 1000 new samples of equal size as the original are drawn with replacement.

finite samples than analytical methods which are based on asymptotic distributions. In addition, bootstrapping is easier to program.

In Appendix 5.9.5, results are also shown for an alternative specification using OLS on an untransformed outcome variable. The advantage is that the coefficients are directly interpretable as marginal effects.

5.5.3 Separating the Effects of PD from Other Regional Differences

The most important challenge is to avoid confounding the effects of PD with other variations between regions. In a first step, the analysis is restricted to four cantons (Argovia, Bern, Lucerne, and Zurich) which are similar in socio-economic status and mentality. This serves to greatly limit unobserved regional variation. The total area selected measures only around 10,585km^2 (which is smaller than Connecticut, one of the smallest states of the US).

In addition, region-specific fixed effects are added to the estimation. The entity here is the 'premium setting region' which is defined by the Swiss federal health authority (FOPH) as having limited within variation in health care expenditures. Insurers are allowed to differentiate premiums between, but not within these regions. The bottom lines of Table 5.1 display the regional distribution of individuals. Even in areas dominated by PD, many individuals use the pharmacy, and vice versa. Therefore, it is possible to include these fixed effects without creating the problem of perfect multi-collinearity with the PD variable.

Admittedly, it is still debatable whether physicians in the PD regions react in the same manner to financial incentives as physicians outside the PD regions. If physicians who are 'strongly attached to money' practice mostly in PD regions, the estimated effect of PD might be biased upward.[11] Unfortunately, a detailed analysis of physicians' choice of location is beyond the scope of this study. Data to conduct such an analysis would be very difficult to obtain for two main reasons. First, anecdotic evidence suggests that a physician's private situation and personal relationship

[11]Longitudinal data is frequently used to control for unobserved heterogeneity. However, this approach could not be followed here because the use of PD is fairly constant over time.

with her predecessor are crucial factors for her choice of location. Second, due to different ages of physicians, either a very long stream of data or surveys on physicians' personal histories would be necessary. It is argued here that the regional fixed effects in combination with the selection of a small area within Switzerland serve to sufficiently attenuate the problem of endogenous choice of location.

5.6 Data

Patient level data of about 240,000 individuals was provided by CSS, the largest Swiss insurer. In addition to sociodemographic variables, information is available on expenditure for four different types of medical care. The variable of main interest - the main supplier of drugs - is represented by a dummy indicating that an individual has purchased 66 or more percent of her drugs from physicians rather than from pharmacists. A sensitivity analysis indicated that the results do not markedly change in response to a variation of the threshold. For individuals with no pharmaceutical spending, the market share of physician-dispensed drugs in their municipalities is used instead of the individual value.

Table 5.1 shows descriptive statistics by sector. Health care expenditure per person is measured with very large standard errors. Given these standard errors, the difference in raw means between the two sectors is far from significant for all types of care. However, mean HCE are lower for individuals who purchase from the physician. Surprisingly, monthly drug expenditure is also about ten percent lower in the physician dispensing group; however the standard errors are again very high. In the case of physician services, descriptive statistics indicate that PD is associated with a higher use of primary care but a lower use of specialists' services.

Now turning to the shares of individuals for whom positive amounts of care are observed, these are higher in the PD sector. It seems that individuals in the PD regions are more likely to visit their doctors at least once within a period of time (1 year). A possible explanation is the increased availability of primary care physicians in the PD regions (see Table 5.4).

Table 5.1: Descriptive Statistics in Non-PD/PD Sector, Year 2005 (* 2007 election).

	Non-PD Sector		PD Sector	
	Mean	Std. Dev.	Mean	Std. Dev.
Consumption of Health Care Services in 2005				
Health Care Expenditure (HCE) per month	287.128	(728.416)	250.919	(654.441)
Drug Expenditure (C_{drug}) per month	64.977	(184.1)	56.086	(211.743)
Primary Care Expenditure (C_{GP})	25.980	(47.337)	29.315	(50.986)
Specialist Expenditure (C_{Spec}) per month	35.743	(90.049)	28.982	(85.438)
Hospital Expenditure (C_{Hosp}) per month	78.964	(476.25)	61.000	(371.839)
Dummy for HCE = 1	0.752	(0.432)	0.795	(0.404)
Dummy for $C_{Drug} = 1$	0.634	(0.482)	0.705	(0.456)
Dummy for $C_{GP} = 1$	0.557	(0.497)	0.665	(0.472)
Dummy for $C_{Spec} = 1$	0.502	(0.500)	0.497	(0.5)
Dummy for $C_{Hosp} = 1$	0.113	(0.316)	0.096	(0.295)
Community Characteristics				
Urban center	0.277	(0.447)	0.140	(0.347)
Suburban municipality	0.341	(0.474)	0.354	(0.478)
Socio-economic status index (community)	51.707	(6.276)	49.985	(7.286)
Share voters conservative (SVP)*	33.360	(8.957)	31.010	(8.830)
Share voters center left (CVP)*	11.495	(7.186)	20.887	(14.331)
Share voters left (SP)*	19.754	(6.409)	14.133	(6.745)
Individual Characteristics				
Age	48.381	(17.822)	48.454	(17.957)
Share Male	0.449	(0.497)	0.464	(0.499)
Deductible Minimum	0.555	(0.497)	0.560	(0.496)
Deductible Medium	0.216	(0.411)	0.215	(0.411)
Deductible High	0.230	(0.421)	0.226	(0.418)
Individual is in one or more PCG	0.225	(0.418)	0.185	(0.388)
Number of individuals in each premium region[12]				
Argovia Region 1	57,215		8,283	
Zurich Region 1	11,491		1,728	
Zurich Region 2	15,146		16,404	
Zurich Region 3	9,009		17,250	
Berne Region 1	9,402		2,560	
Berne Region 2	10,845		6,925	
Berne Region 3	2,465		3,359	
Lucerne Region 1	2,207		23,310	
Lucerne Region 2	956		16,538	
Lucerne Region 3	1,435		26,263	
Observations	120,193		122,620	

PD = 1: The individual has bought at least 66 percent of her drugs from a physician.

Considering the explanatory variables, individuals who buy at a pharmacy are more likely to live in urban or suburban areas, in comparison with individuals who

buy their drugs from physicians. This is intuitive given the higher availability of pharmacies in urban areas. Socio-economic status is slightly higher among individuals who buy drugs at the pharmacy. Differences in mentalities among communities are captured by the share of voters for three of the four largest parties in the parliamentary election of 2007. Table 5.1 shows that PD-patients are more likely than non-PD patients to live in communities with strong support for CVP, a center-left party with catholic roots. The leftist, social-democrat SP is stronger in the non-PD areas, while the conservative, right-wing SVP is quite strong in both sectors. One large party, the center-right, pro-market FDP, has been excluded in order to avoid multi-collinearity.

In terms of personal characteristics, individuals in both sectors are quite similar. The average age is roughly 48 in both sectors and the share of males is below 50 percent. The deductible is the amount of health care expenditure that an individual has to pay out of pocket before the insurance kicks in. Chosen by the patient, it is a good indicator for self-assessed health status. Table 5.1 shows no difference between individuals who choose different drug providers. The presence of chronic illness is described by 21 pharmaceutical cost groups.[13] These groups are considerably more frequent in the pharmacy sector, which probably reflects better reporting by pharmacies rather than their true prevalence.[14]

5.7 Results

The testing of Hypotheses (1) to (3) is discussed in Sections (5.7.1) to (5.7.3). Table 5.2 displays the results of both parts of the two-part model. For each type of health care expenditure (ie. drugs, GP care, specialist care, and hospital inpatient care) the respective left column contains the marginal effects from a probit estimation

[13]The definition of these groups is similar to the ones described by Lamers and Van Vliet (2003).

[14]Pharmacies commonly bill insurers directly by electronic transmission. In contrast, some physicians choose to bill the patients, who then send the bills to insurers for reimbursement. As the direct-to-patient bill is not transmitted electronically, detailed drug information is often lost. Therefore, the PCG cannot always be determined.

of the probability of observing this type of expenditure. The marginal effects are evaluated at the median of the explanatory variables.[15] For dummy variables that describe mutually exclusive groups (such as deductible levels or community types), the most frequent group is set to one while all others are set to zero. In the case of interactions, the marginal effects are calculated according to the formulas provided by Ai and Norton (2003) (see Appendix 5.9.3). The respective right column displays results of a GLM estimation of the log drug expenditure, conditional on observing a positive amount of drugs.

5.7.1 Hypothesis 1: Physician Dispensing and Drug Expenditure

The coefficient of main interest pertains to the fact that an individual buys most her drugs from a physician. It increases the probability of buying drugs by almost 16 percentage points, speaking strongly in favor of Hypothesis (1). Interestingly, results in the second column show that once drug use is initiated, the expenditure level is markedly lower in the PD sector. This effect can be caused either by lower drug quantities (thereby contradicting Hypothesis 1), or cheaper drugs being prescribed by dispensing physicians. The latter is in line with findings presented in Chapter 4 of this thesis that dispensing physicians sell significantly more generic drugs than pharmacies do.

[15] For dummy variables, the marginal effects reflect a change from 0 to 1.

Table 5.2: Two-part Estimation of Expenditure on Drugs, General Practitioners, Specialists, and Hospitals.

	$Pr(C_{Drug} > 0)$ Marg. Effect	$C_{Drug}\|C_{Drug} > 0^a$ Coefficient	$Pr(C_{GP} > 0)$ Marg. Effect	$C_{GP}\|C_{GP} > 0^a$ Coefficient	$Pr(C_{Spe} > 0)$ Marg. Effect	$C_{Spe}\|C_{Spe} > 0^a$ Coefficient	$Pr(C_{Hos} > 0)$ Marg. Effect	$C_{Hos}\|C_{Hos} > 0^a$ Coefficient
Buys from physician	0.159***	-0.168***	0.162***	0.032***	0.041***	-0.133***	-0.023***	-0.059*
	(0.004)	(0.023)	(0.003)	(0.009)	(0.003)	(0.013)	(0.003)	(0.030)
Buys from phy. * DEDmedium	0.019***	-0.082*	0.010**	-0.019	0.017**	0.020	0.001	-0.122*
	(0.005)	(0.039)	(0.004)	(0.014)	(0.005)	(0.022)	(0.003)	(0.051)
Buys from phy. * DEDhigh	0.063***	0.072	0.088***	-0.048**	0.078***	0.104***	0.013***	0.030
	(0.006)	(0.049)	(0.005)	(0.018)	(0.005)	(0.027)	(0.003)	(0.070)
DEDmedium	-0.090***	-0.240***	-0.061***	-0.084***	-0.062***	-0.085***	-0.014***	-0.003
	(0.004)	(0.029)	(0.004)	(0.011)	(0.004)	(0.016)	(0.003)	(0.035)
DEDhigh	-0.315***	-0.778***	-0.286***	-0.263***	-0.265***	-0.390***	-0.070***	-0.155**
	(0.004)	(0.037)	(0.004)	(0.014)	(0.004)	(0.020)	(0.003)	(0.049)
Male	-0.166***	0.063	-0.264***	-0.799***	-0.355***	0.236**	-0.137***	0.127
	(0.017)	(0.125)	(0.015)	(0.046)	(0.012)	(0.075)	(0.005)	(0.176)
Age	0.002***	0.041***	-0.001	-0.000	0.006***	0.028***	-0.006***	0.015***
	(0.000)	(0.003)	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.004)
Age squared	0.001*	-0.253***	-0.001***	0.059***	0.006***	-0.287***	-0.005***	-0.080*
	(0.000)	(0.028)	(0.000)	(0.011)	(0.000)	(0.017)	(0.002)	(0.035)
Male * age	-0.114***	0.007	-0.158***	0.018***	-0.156***	-0.007*	-0.033***	0.004
	(0.002)	(0.005)	(0.002)	(0.002)	(0.002)	(0.003)	(0.001)	(0.007)
Male * age squared	-0.171***	-0.094*	-0.269***	-0.104***	-0.222***	0.079**	-0.13***	-0.049
	(0.017)	(0.046)	(0.016)	(0.017)	(0.006)	(0.028)	(0.008)	(0.059)
Suburban community	0.009	0.036	0.014**	0.025*	0.005	0.001	0.007*	0.062
	(0.005)	(0.032)	(0.004)	(0.012)	(0.004)	(0.018)	(0.003)	(0.042)
Affluent community	0.052***	0.020	0.014	-0.051*	0.013	0.028	0.011	-0.048
	(0.008)	(0.065)	(0.008)	(0.024)	(0.009)	(0.036)	(0.007)	(0.083)
Rural & commuter community	-0.016**	-0.014	-0.007	-0.038*	-0.005	-0.095***	0.006	-0.000
	(0.006)	(0.043)	(0.005)	(0.016)	(0.006)	(0.024)	(0.005)	(0.055)
Share voters conservative (SVP)	0.000	0.002	-0.000	0.000	0.000	0.000	-0.001*	-0.003
	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)
Share voters left (SP)	0.001*	-0.002	0.001**	0.001	-0.001***	-0.005***	-0.000	-0.007*
	(0.000)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.003)
Share voters centre left (CVP)	0.002***	-0.002	0.002***	-0.001	0.002***	-0.002	-0.001**	-0.006
	(0.000)	(0.003)	(0.000)	(0.001)	(0.000)	(0.002)	(0.000)	(0.004)
N	242,813	159,371	242,813	148,429	242,813	121,316	242,813	25,330
AIC	246,449	1,645,069	278,195	1,402,121	299,386	1,236,954	150,369	377,217

Additional regressors: 10 Premium Region Fixed Effects, 21 Pharmaceutical Cost Groups, 4 Additional Community Types, Population of the community, Socio-economic status index.

Standard errors in parentheses, *p<0.05,** p<0.01, ***p<0.001, ^{a} in logs

It is often argued by pharmacists and their lobbying groups that financial incentives pertaining to PD weaken physician agency for the patient. Patients with high deductibles have an increased interest in keeping their health care expenditure low. Therefore, the interaction between PD and a high deductible (Buys from physician · DED high) can serve as a test of physician agency.[16] If the interaction between PD and the deductible level is significantly positive, PD providers are less responsive to patient interest, pointing to imperfect agency. The results in the first column of Table (5.2) confirm this notion. Patients with high deductibles have a higher probability of drug use in the PD than in the non-PD sector. However, this effect may be caused by the compliance behavior of patients rather than the prescribing behavior of physicians. It is also likely that individuals in the pharmacy sector often do not buy the prescribed drugs, while patients in the PD sector feel more urged to do so. Once drug expenditure are initiated, the interaction of a medium deductible with PD is negative, contradicting lower agency. The interaction of the high deductible is insignificant. Overall, the evidence concerning agency is inconclusive.

5.7.2 Hypothesis 2: Physician Dispensing and Primary Care Services

Hypothesis 2 states that PD reduces the amount spent on physician treatments because physicians reallocate their resources to drug prescriptions. Surprisingly, the opposite association is observed in the case of primary care services. As shown in Column (3) and (4) of Table 5.2, patients who buy drugs from physicians have a markedly higher probability of primary care expenditure and also a larger expected amount of expenditure than individuals who buy from pharmacies.

A possible explanation for the theoretical model's failure to predict the increase in spending for primary care is that it focuses on only one period. In a more dynamic setting, providing physicians with dispensing rights is likely to increase the supply

[16]It is important to note that the coefficient of the deductible level cannot be interpreted directly because it is composed of two components, the 'self-selection effect' (healthier individuals are more likely to choose a high deductible) and the 'incentive effect' (higher deductibles reduce moral hazard). For an econometric separation of the two effects see Van Kleef et al. (2008), Gardiol et al. (2005), and Chapter 3 of this thesis.

of primary care physicians. In fact, by reducing the income gap between them and better paid specialists, dispensing might motivate more young physicians to involve themselves in primary care supply. Furthermore, those regions that allow dispensing might attract general practitioners from other regions. The numbers displayed in Table 5.4 in the appendix point in this direction because the PD regions have a greater share of physicians in primary care when compared to the Non-PD regions. This finding is potentially interesting for countries such as the US who are experiencing a shortage of primary care physicians [Goroll et al. (2007)]. However, an analysis of the dynamics of physicians' choice of field and location is beyond the scope of this text.

The interaction term of buying from physicians and having a high deductible indicates that the likelihood of positive primary care expenditure for individuals with high deductibles is greater in the PD than in the pharmacy sector. However, this cannot be attributed to physician behavior because the decision of whether or not to visit a physician is taken by the individuals themselves. On the second stage of the model, the interaction term is negative and does not confirm the notion of reduced agency by dispensing physicians.

5.7.3 Hypothesis 3: Physician Dispensing and Referrals

Hypothesis 3 states that physician dispensing reduces the number of referrals because the treatment of patients with severe illness is more profitable to primary care physicians. Table 5.2 provides some evidence for this hypothesis, as the association of buying drugs from physicians with hospital expenditure is negative in both parts of the two-part model.

The case of specialist physicians is more complicated, as patients in the dataset may visit specialists without referrals. Frequently visited specialists such as ophthalmologists or gynecologists are usually seen without prior consultation by a primary care physician, while 'highly specialized' specialists such as endocrinologists are not. Table 5.2 shows a positive association of PD and specialists' expenditure on the first

part of the two-part model. This is likely being driven by the frequently visited specialists who show the same patterns as primary care physicians. By way of contrast, the second part of the two-part model is likely being driven by the 'specialized specialists', whose services are often related to chronic illness. Here, the association with PD is negative, indicating that primary care physicians in the PD sector provide a larger share of the care for the chronically ill than in the non-PD sector. It is intuitive that the increased income from PD motivates primary care physicians to expand their knowledge in relation to chronic illness.

5.7.4 Physician Dispensing and Health Care Expenditure

For policy, the impact of a piece of regulation on actual expenditure is often of greater interest than the relative impact. To evaluate this, Table 5.3 shows the combined incremental effect calculated according to Equation (5.13). In the first bloc, the effect of buying from a physician on drug expenditure per month is displayed. The mean reduction is statistically significant, but of limited size (CHF - 5 or EUR - 3.33 per patient and month). Contrary to the expectations of economists, drug dispensing by physicians does not increase drug expenditure in our data.

In the case of primary care expenditures, the incremental effect is significantly positive (roughly CHF + 7 per patient and month), pointing at an increased utilization by individuals who buy their drugs from physicians. For specialist physicians, the incremental effect is insignificant, while hospital expenditures seem to be lower for individuals who choose to buy their drugs from physicians than for other individuals.

Last but not least, results are discussed for health care expenditure in general. For brevity, the results of the two-part model are shown in the appendix (Table 5.6). Physician dispensing increases the likelihood of positive health care expenditure. This is unsurprising given our previous finding that PD increases the use of primary care and drugs. These two types of care are frequently used for relatively minor ailments and are most likely to be substituted by self medication (no formal care). However, physician dispensing is associated with a sizeable decrease of the amount of care that

Table 5.3: Combined Incremental Effects and Their Bootstrapped Standard Errors.

	Incremental Eff. in CHF	Bootstrap Std. Err.	Bootstrap 5th Percentile	Bootstrap 95th Percentile
Drug Expenditure				
Buys from physician	-5.357	(1.461)	-7.799	-3.010
Buys from physician * DEDmedium	-0.371	(1.210)	-1.546	2.379
Buys from physician * DEDhigh	5.958	(1.507)	3.491	8.348
Primary Care				
Buys from physician	6.736	(0.231)	6.364	7.132
Buys from physician * DEDmedium	-0.691	(0.201)	-1.028	-0.374
Buys from physician * DEDhigh	-1.092	(0.218)	-1.445	-0.729
Specialists				
Buys from physician	-0.861	(0.377)	-1.631	-0.328
Buys from physician * DEDmedium	1.326	(0.384)	0.711	2.055
Buys from physician * DEDhigh	4.031	(0.441)	3.426	4.868
Hospitals				
Buys from physician	-15.439	(2.067)	-18.894	-12.049
Buys from physician * DEDmedium	-2.498	(2.232)	-6.175	1.147
Buys from physician * DEDhigh	9.992	(2.621)	5.687	14.481
Health Care Expenditure				
Buys from physician	-14.858	(3.826)	-21.280	-8.662
Buys from physician * DEDmedium	0.663	(3.411)	-4.960	6.359
Buys from physician * DEDhigh	25.903	(4.287)	18.863	32.874
Observations	1,000	1,000	1,000	1,000

1 CHF $\approx$ 0.66 EUR at 2005 exchange rates.

is used. This could be a mixed effect of the use of cheaper generic drugs and fewer referrals thanks to more primary care. Many authors, notably in the USA, argue that an increased use of primary care generates cost savings in other types of care [Goroll et al. (2007), Bodenheimer (2006)].

The combined incremental effect (Table 5.3) is about negative 15 CHF per month, with the 95 percent confidence interval reaching from CHF - 21 to - 8 CHF. It indicates that PD leads on average to lower health care expenditure in these Swiss cantons, but the effect is not particularly large. From these results, it seems unlikely that outlawing physician dispensing, as it is proposed by pharmacists and their lobbying groups, would markedly influence health care expenditure per capita in Switzerland.

These findings contradict opinions stated by many physicians as well as opponent opinions stated by pharmacists and their lobbyists. Physicians often argue that they are motivated intrinsically and are completely immune to financial incentives. This is not confirmed because the PD variable is significant in most estimates. On the other hand, pharmacists often argue that financial incentives are so influential that the income earned by drug dispensing necessarily leads to higher health care expenditure per capita. This has also not been confirmed.

5.7.5 Robustness Check

In order to test the robustness of our results against the econometric specification, Table 5.7 in the appendix shows the results of an OLS regression applied to the untransformed data. The advantage of this specification is that the coefficients can be directly interpreted as marginal effects. OLS yields consistent estimates of the coefficients if the number of observations is large enough. The standard errors need to be adjusted for heteroscedasticity.

Except for drug expenditure, the OLS coefficients have the same sign and roughly the same magnitude as the average marginal effects estimated by the 2PM. For drug expenditure, the effect of PD switches from being around CHF - 5.35 to being around CHF + 3.70. Therefore, the OLS estimates confirm Hypothesis (1) stating that drug dispensing by physicians increases drug expenditures. However, the estimated effect is still quite small, and thus its economic relevance is limited.

5.8 Conclusions

Markets for medical care are strictly regulated even in nations that claim to be market economies. The reason is the information advantage of medical professionals vis-a-vis patients and insurers. This paper analyzes the case of drug prescribers. Most OECD countries prevent prescribers of drugs from dispensing, fearing that financial incentives will lead to inefficiently high drug expenditure. Exceptions are countries in Southeast Asia where the physician and the pharmacist were traditionally the

same person. Switzerland is one of the few countries where both arrangements are observed, which enables comparison of expenditure in both settings. While several previous studies have examined this subject, their conclusions point in different directions. Unfortunately, no study has yet analyzed the differences in health outcomes, a limitation which is also true for this research.

From a theoretical model of physician decision making, three main hypotheses are derived. Hypothesis (1) states that allowing prescribers to dispense on their own account is associated with increased quantities of drugs being prescribed. Hypothesis (2) states that allowing prescribers to dispense reduces their incentive to provide treatments because some treatments can be substituted by more profitable drugs. Hypothesis (3) states that due to higher incomes earned per case, dispensing prescribers have less incentives to refer patients to hospitals and specialists.

These three hypotheses are tested using patient-level data of over 240,000 Swiss patients. It has been found that patients who buy most of their drugs from physicians have a higher likelihood of using drugs than patients who buy at pharmacies, giving support to Hypothesis (1). However, if positive drug expenditure is observed, this will be lower for patients who buy from physicians. Combining the two effects results in a moderately negative association of physician dispensing with drug expenditure. Therefore, this research fails to confirm increased drug spending by patients who buy drugs from physicians, which is at odds with microeconomic theory. Hypothesis (2) must also be rejected because patients who buy drugs from physicians consume more primary care services than patients who buy drugs from pharmacies. Hypothesis (3) is confirmed by the fact that patients who buy drugs at the physician's office use lower hospital services on average than patients who buy drugs at the pharmacy.

Turning to total health care expenditure, this research finds a negative association between physicians selling drugs and total health care expenditure per capita, which is statistically significant but only of a moderate size. Therefore, it fails to provide supporting evidence for the claim that banning physicians from dispensing drugs would markedly reduce health care expenditure. In this situation, it might be

sensible to liberalize the separation between prescribers and sellers of drugs (either by allowing pharmacists prescribe or physicians dispense a selection of drugs). In this way, competition among drug providers is increased, which enhances responsiveness to patient preferences and decreases drug mark-ups (unless prevented by regulation). Moreover, it could be used as a means to increase the supply of primary care providers in regions where these are at a premium.

Acknowledgments

This paper was partly written during my visit at Boston University. Helpful suggestions by Randall Ellis, and financial support of the Swiss National Science Foundation (SNF) are gratefully acknowledged. In addition, I would like to thank the CSS Institute for Empirical Health Economics for providing data and many insights, Ilja Neustadt, Peter Zweifel, Albert Ma, Philipp Morf, Maurus Rischatsch, Giovanni Mastrobuoni, Konstantin Beck, and participants at the 8th World Congress on Health Economics in Toronto for helpful discussion and comments.

5.9 Appendix

5.9.1 Regulation of Physician Dispensing in the Swiss Cantons

Table 5.4: Descriptive Statistics per Canton ('State'), year 2006.

Cantons	Language	Regulation on PD	HCE 2006	Drug Cost 2006	% Drug Turnover sold by physicians	Physician density[1]	% GPs	Pharmacy density[2]
AI	G	1	1,821	347	0.86	1.19	0.33	6.61
NW	G	1	2,057	398	0.79	1.14	0.36	5.08
AR	G	1	2,149	395	0.83	1.87	0.4	11.48
OW	G	1	2,176	411	0.86	1.17	0.56	11.97
UR	G	1	2,191	407	0.89	1.19	0.44	5.78
ZG	G	1	2,233	410	0.74	1.82	0.33	12.11
LU	G	1	2,287	443	0.82	1.49	0.37	9.49
SZ	G	1	2,306	448	0.79	1.35	0.43	8.69
GL	G	1	2,321	469	0.86	1.45	0.51	5.26
SG	G	1	2,335	466	0.8	1.63	0.38	10.58
TG	G	1	2,355	402	0.77	1.34	0.42	9.75
GR	G/I	2	2,398	498	0.39	1.75	0.41	20.37
VS	F/G	2	2,463	568	0.05	1.75	0.33	37.5
AG	G	2	2,484	541	0.14	1.51	0.32	19.03
FR	F/G	2	2,572	551	0.08	1.55	0.32	25.02
SH	G	2	2,611	567	0.35	1.99	0.4	17.49
SO	G	1	2,691	552	0.65	1.65	0.39	10.5
JU	F	3	2,713	631	0.07	1.69	0.35	27.93
ZH	G	2	2,725	528	0.43	2.31	0.29	16.91
BL	G	1	2,856	565	0.52	2.01	0.31	15.05
NE	F	3	2,951	710	0.07	2.08	0.26	33.65
BE	G/F	2	2,984	560	0.42	2.13	0.3	19.34
VD	F	3	3,144	682	0.03	2.46	0.25	35.74
TI	I	3	3,287	707	0.06	2.04	0.27	54.8
BS	*G*	*3*	*3,585*	*751*	*0.09*	*3.81*	*0.18*	*36.25*
GE	*F*	*3*	*3,625*	*773*	*0.04*	*3.29*	*0.17*	*38.88*

Languages: G=German, I=Italian, F=French

Regulation of Dispensing: 1=PD allowed to all physicians, 2= Mixed Systems, 3=PD not allowed

BS and *GE* are 'city states' and more urban than the other cantons.

LU and AG are interesting because they are similar in urbanisation, but have differing laws on PD.

[1] Number of physicians per 1,000 population, [2] Number of pharmacies per 100,000 population

5.9.2 Testing Multi-collinearity in the Sample Selection Model

The variance inflation factor (VIF) is used to test the severity of the multi-collinearity problem in the SSM. The outcome variables are transformed by the log before applying the standard SSM. Because of high collinearity of age and age squared as well as the interacted gender variables, the mean VIFs are calculated excluding the age and

gender variables. For comparison, the mean VIF is also shown for an OLS regression excluding inverse mills ratio. Table 5.5 clearly supports the notion by Leung and Yu (1996) that the multi-collinearity problem is severe if censoring is high. For rarely used types of care such as specialized physicians or hospitals, the SSM is not applicable unless there are exclusion restrictions. Although the VIF of the inverse Mills ratio is high throughout, the SSM is a valuable alternative if censoring has a weak effect.

Table 5.5: Test for Multi-collinearity Based on the VIF.

	HCE	Drugs	Primary Care	Specialized Physicians	Hospitals
VIF inverse mill ratio	9.07	9.33	36.83	68.53	999.94
Mean VIF SSM	2.35	2.33	3.42	4.17	29.96
Mean VIF OLS	2.15	2.13	2.17	2.13	2.31

5.9.3 Interaction Effects in Probit Models

In nonlinear models such as the probit, the marginal effect of changing two interacted variables is not equal to the marginal effect of just the interaction term [Norton et al. (2004)]. Instead, the marginal effect is found by taking the cross derivative, if the two interacted variables are continuous, or the double discrete difference, if they are dummies. For example, let x_1 and x_2 be two interacted dummy variables, X a vector of additional independent variables, $u = \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + X\beta$, and $F(u)$ a continuous and differentiable function. The double discrete difference reads

$$\begin{aligned} \frac{\Delta^2 F(u)}{\Delta x_1 \Delta x_2} &= \frac{\Delta\{F(\beta_1 + \beta_2 x_2 + \beta_{12} x_2 + X\beta) - F(\beta_2 x_2 + X\beta)\}}{\Delta x_2} \\ &= F(\beta_1 + \beta_2 + \beta_{12} + X\beta) - F(\beta_2 + X\beta) \\ &\quad -F(\beta_1 + X\beta) + F(X\beta) \end{aligned} \tag{5.15}$$

The standard errors of the marginal effects can be approximated using the delta method [Cameron and Trivedi (2005)]. Let $\mathrm{g}(\hat{\beta})$ be a column vector of marginal

effects. For example, if there are two interacted dummy variables and one additional control variable (x_i), $\mathbf{g}(\hat{\beta})$ reads $(\frac{\Delta F(u)}{\Delta x_1}, \frac{\Delta F(u)}{\Delta x_2}, \frac{\Delta^2 F(u)}{\Delta x_1 \Delta x_2}, \frac{\Delta F(u)}{\Delta x_i})$. Now, let there be a matrix ($\mathbf{G}(\hat{\beta})$) of partial derivatives of $\partial \mathbf{g}(\hat{\beta}) / \partial \beta'$. In the aforementioned example of two interacted dummy variables and one additional control variable, $\mathbf{G}(\hat{\beta})$ reads

$$\mathbf{G}(\hat{\beta}) = \begin{bmatrix} \frac{\partial(\frac{\Delta F(u)}{\Delta x_1})}{\partial \beta_1} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_1})}{\partial \beta_2} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_1})}{\partial \beta_{12}} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_1})}{\partial \beta_i} \\ \frac{\partial(\frac{\Delta F(u)}{\Delta x_2})}{\partial \beta_1} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_2})}{\partial \beta_2} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_2})}{\partial \beta_{12}} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_2})}{\partial \beta_i} \\ \frac{\partial(\frac{\Delta^2 F(u)}{\Delta x_1 x_2})}{\partial \beta_1} & \frac{\partial(\frac{\Delta^2 F(u)}{\Delta x_1 x_2})}{\partial \beta_2} & \frac{\partial(\frac{\Delta^2 F(u)}{\Delta x_1 x_2})}{\partial \beta_{12}} & \frac{\partial(\frac{\Delta^2 F(u)}{\Delta x_1 x_2})}{\partial \beta_i} \\ \frac{\partial(\frac{\Delta F(u)}{\Delta x_i})}{\partial \beta_1} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_i})}{\partial \beta_2} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_i})}{\partial \beta_{12}} & \frac{\partial(\frac{\Delta F(u)}{\Delta x_i})}{\partial \beta_i} \end{bmatrix}$$

With $\hat{\mathbf{V}}(\beta)$ denoting the estimated variance-covariance matrix of the probit estimation, the variance of the marginal effects $\hat{\mathbf{V}}(\mathbf{g}(\hat{\beta}))$ is obtained by,

$$\hat{\mathbf{V}}(\mathbf{g}(\hat{\beta})) = \mathbf{G}(\hat{\beta}) \hat{\mathbf{V}}(\beta) \mathbf{G}(\hat{\beta})' \tag{5.16}$$

5.9.4 Results for Health Care Expenditure

Table 5.6: Two-Part Results for Health Care Expenditure.

	Pr(HCE >0) Marg. Effect	HCE|HCE> 0 Coefficients
Buys from physician	0.111***	-0.138***
	(0.003)	(0.017)
Buys from physician * DEDmedium	0.018***	-0.063*
	(0.003)	(0.028)
Buys from physician * DEDhigh	0.074***	0.112***
	(0.004)	(0.031)
DED Medium	-0.067***	-0.175***
	(0.004)	(0.020)
DED High	-0.243***	-0.662***
	(0.004)	(0.023)
Male	-0.318***	-1.017***
	(0.019)	(0.085)
Age	-0.007***	-0.016***
	(0.000)	(0.002)
Age squared	-0.005***	0.272***
	(0.000)	(0.020)
Male * age	-0.119*	0.036***
	(0.002)	(0.003)
Male * age squared	-0.231***	-0.280***
	(0.021)	(0.032)
Suburban community	0.010*	0.044
	(0.004)	(0.023)
Affluent community	0.024***	0.020
	(0.007)	(0.046)
Suburban - mixed community	-0.006	-0.024
	(0.005)	(0.030)
Industrial community	-0.009*	0.050
	(0.005)	(0.029)
Rural & commuter community	-0.002	-0.003
	(0.005)	(0.030)
AIC	210,527	2,471,028
N	242,641	187,821

Additional regressors are the same as those in Table 5.2.

Standard Errors in Parentheses,*p<0.05,** p<0.01,***p<0.001

5.9.5 Robustness Check Using OLS

Table 5.7: OLS Results (Robustness Check).

	Drug	Primary Care	Specialists	Hospitals	HCE
Buys from physician	3.742*	8.326***	-2.547***	-16.694***	-8.363
	(1.645)	(0.332)	(0.701)	(3.101)	(5.012)
Buys from physician * DEDmedium	0.512	-1.305**	3.121***	-3.425	1.168
	(1.782)	(0.473)	(0.907)	(4.571)	(6.619)
Buys from physician * DEDhigh	4.220**	-1.389***	9.237***	17.069***	38.831***
	(1.289)	(0.355)	(0.639)	(3.440)	(4.945)
DED Medium	-14.881***	-3.514***	-6.414***	-4.531	-50.624***
	(1.116)	(0.324)	(0.649)	(3.873)	(5.217)
DED High	-28.937***	-11.850***	-22.887***	-29.822***	-126.526***
	(0.801)	(0.255)	(0.495)	(2.433)	(3.487)
Male	0.293	-10.449***	-4.652*	-26.313*	-239.739***
	(4.016)	(1.286)	(2.118)	(11.567)	(18.451)
Age	0.907***	-0.333***	1.346***	-3.007***	-19.659***
	(0.105)	(0.043)	(0.065)	(0.354)	(0.599)
Age squared	-3.616***	5.908***	-14.688***	35.550***	243.562***
	(1.075)	(0.448)	(0.646)	(3.590)	(6.352)
Male * age	-0.014	-0.161**	-0.463***	-0.141	8.639***
	(0.184)	(0.061)	(0.096)	(0.561)	(0.885)
Male * age squared	1.679	3.553***	8.188***	14.893*	-72.858***
	(1.850)	(0.651)	(0.985)	(6.126)	(9.576)
Suburban community	1.714	1.247***	0.289	11.526**	19.302***
	(1.471)	(0.331)	(0.641)	(3.739)	(5.327)
Affluent community	4.934	-0.655	4.834**	4.053	14.637
	(3.263)	(0.766)	(1.613)	(7.394)	(10.962)
Suburban - mixed community	-1.145	0.402	-3.979***	4.227	3.954
	(1.861)	(0.452)	(0.796)	(4.345)	(6.467)
Industrial community	1.853	-0.244	-3.104***	12.475**	25.141***
	(1.678)	(0.389)	(0.705)	(4.536)	(6.446)
Rural & commuter community	1.340	-0.930*	-2.735***	8.064	11.934
	(1.857)	(0.405)	(0.802)	(4.294)	(6.301)
Share voters conservative (SVP)	-0.013	-0.000	-0.026	-0.507*	-0.790*
	(0.117)	(0.022)	(0.040)	(0.218)	(0.328)
Share voters left (SP)	-0.031	0.055**	-0.117***	-0.584***	-0.714**
	(0.092)	(0.019)	(0.030)	(0.177)	(0.274)
Share voters centre left (CVP)	0.088	0.018	-0.002	-1.236***	-1.383**
	(0.158)	(0.032)	(0.061)	(0.331)	(0.500)
Socio-economic status	0.147	0.174***	0.400***	0.222	0.995**
	(0.089)	(0.024)	(0.039)	(0.211)	(0.313)
R^2	0.247	0.168	0.083	0.033	0.185
N	242,813	242,813	242,813	242,813	242,813

Additional regressors: 10 Premium Region Fixed Effects, 21 Pharmaceutical Cost Groups, 2 Additional Community Types, Population of the Community, Standard Errors in parentheses, *p<0.05,** p<0.01,***p<0.001.

Chapter 6

Conclusion

This section summarizes the main policy implications of the thesis and discusses possible extensions of each essay.

This thesis analyzes different examples of how market mechanisms can be useful to allocate resources in health care markets – in spite of considerable information asymmetries. However, if markets are regulated with the aim of achieving social goals, regulation has to be carefully designed in order to avoid giving unintended incentives to market players. Results of Chapter 2 demonstrate that community rating on its own cannot ensure equal access to chronically ill individuals in health insurance markets. In order to mitigate the insurers' incentives to 'chase the low risks', community rating has to be combined with health-based risk adjustment. Although such schemes can never be perfect because the regulator has an information disadvantage vis-à-vis health insurers, results in Chapter 2 indicate that effective formulas can be designed using readily available health indicators.

Policy implication 1: If community rating is imposed in health insurance markets, health-based risk adjustment is a necessity.

Designing an effective risk adjustment scheme is challenging if insurers are granted the freedom to offer plans with different coverage designs. Empirical evidence from Chapter 3 suggests that both supply-side and demand-side cost sharing plans attract mainly healthy individuals. However, both options also achieve marked reductions of

moral hazard. Cost reductions due to reductions of moral hazard ('true' cost reductions) that are attained by such plans depend on the expenditure level of the group who chooses the respective plan because healthy individuals have lower potential for cost savings than sick individuals. The question, then, is to what extent insurers should be allowed to pass on gross savings to consumers. It seems undisputed that the 'true' cost savings should be passed on. Indeed, if 'true' cost savings are punished by higher contributions to the risk adjustment scheme, incentives for offering such contracts are reduced [Lehmann and Zweifel (2004)]. Currently, these considerations are not part of the political debate on risk adjustment. However, they should be taken into consideration in case future amendments of the risk adjustment scheme are envisaged.

Policy implication 2: If contract differentiation is possible, the risk adjustment formula should account for plan-specific cost savings.

It was pointed out in Chapter 3 that both types of cost sharing plans disproportionately attract healthy individuals. However, the selection effects are stronger in the case of demand-side cost sharing. It might be undesired that cost sharing plans are chosen by very healthy individuals only, due to their limited potential for cost reductions. Results in Chapter 3 imply that if insurers wish to attain a favorable ratio of risk-selection and moral hazard effects, promoting supply-side cost sharing seems to be a more promising strategy than promoting demand-side cost sharing.

Policy implication 3: If the ratio between risk-selection and moral hazard effects is taken as a criterion, supply-side cost sharing is more effective than demand-side cost sharing.

The estimations in Chapter 4 imply that a change in the coinsurance rate from 10 to 20 percent for brand-name drugs was very effective in promoting generic substitution. The magnitude of the effect is surprising given that the average additional cost per patient was low compared to income. It seems that physicians and patients

reacted to the fact that they had to pay 'additional' copayment, without necessarily considering its amount. Furthermore, the increase in the coinsurance rate for brand-name drugs likely was interpreted as government advice to use the generic alternative, helping to convince patients and physicians of its safety and quality. This experience should be kept in mind for other instances when treatment alternatives with little difference in documented effectiveness but considerable difference in prices are available. A hefty increase in the rate of coinsurance for the more expensive treatment is sure to meet with resistance from patient lobbying groups and leftist political parties. However, small increases in the coinsurance rate might be quite effective and easier to implement.

Policy implication 4: Introducing treatment-specific rates of coinsurance can successfully promote cost-effective treatment modes even if the additional burden for the patient is limited.

Chapter 5 analyzes the cost implications of giving physicians the right to dispense drugs on their own account. The analysis does not only include expenditure on drugs, but also expenditure on services provided by general practitioners (GPs), specialists, and hospitals. It is found that, although drug dispensing does not directly influence reimbursement of other types of care, quantities of such services are markedly affected. In particular, drug dispensing by physicians is associated with an increased use of GP services but a decreased probability of using hospital services. Therefore, if regulators wish to reduce hospitalizations, a redesign of the reimbursement schemes for GPs could be envisaged.

Policy implication 5: Increasing general practitioners' earnings per case may serve to reduce referrals to hospitals.

This thesis could be a starting point for various extensions. To start with empirical issues, Chapter 2 contains an analysis of insurers' incentives to select risks, assuming a mid-term planning horizon. Using even more years of data, it would become possible to assess to what extent risk profiles are equalized 'by nature' in the long-run

(meaning that some of the ill become healthy and vice versa). Should the 'natural' risk equalization be strong in the long run, risk selection problems could be mitigated by increasing contracting periods (as long as consumers accept to be 'locked into' their insurance contracts).

Chapter 3 demonstrates that supply-side cost sharing is associated with marked cost reductions. However, these results have limited generality unless the mechanism behind the cost reductions are identified. To this end, a deeper analysis of the internal incentive structure of the independent practice association (IPA) would be required. Furthermore, analyzing cost savings by subgroups of patients could be instructive. In particular, the deceased would make an interesting study group. Werblow et al. (2007) found that spending patterns of the deceased differ from those of survivors. Unfortunately, the number of deceased in the IPA plans in our data set was too small for such an analysis. Another interesting study group are patients with specific chronic conditions. More often than not, patients who suffer from chronic conditions are treated by a number of different health care providers. These patients might strongly benefit from increased coordinating effort by their general practitioner. Such an analysis would allow to derive practical guidelines on how efficiency in health care delivery can be improved. According to Beck et al. (2009), the Swiss market currently lacks such guidelines.

Chapter 4 and Chapter 5 consider physicians' choice of location exogenous, with regional level fixed effects controlling for differences between regions. This is justified as long as the physicians who practice in PD regions do not react to financial incentives in a systematically different manner than physicians in non-PD regions do. If physicians who have a 'high valuation for personal income' systematically select into the physician dispensing regions, the estimated effects of PD are possibly overrated. By analyzing physicians' choice of location, it would become possible to separate self-selection effects from the 'true' effects of dispensing. Unfortunately, data to analyze physicians' choice of location does not (yet) exist in Switzerland.

Another extension of Chapters 2, 4, and 5 is to account for competition between insurers, or physicians respectively. In Chapter 2, the cost of risk selection is not explicitly modeled. In order to estimate it, the strategic behavior of competing insurers would have to be taken into account because the cost of attracting low risks increases if many insurers compete for them (provided that the share of low risks in the population is not augmented by risk selection, which is certainly true in the short run). In Chapters 4 and 5, the patient passively accepts the treatment decisions taken by the physician on his behalf. A more realistic model would consider a bargaining process between the two, as proposed by Ellis and McGuire (1990). Furthermore, increased competition likely puts pressure on physicians to respond to patients' preferences.

The results of Chapters 2 and 3 lead to policy implication 2, suggesting that plan-specific cost savings should be accounted for in the risk adjustment formula in order to avoid the dilution of incentives for efficiency. However, implementing this idea is tricky and no 'best practice' approach has yet emerged. On the contrary, Schokkaert and Van de Voorde (2009) showed that if the health care expenditure function is not additively separable into morbidity and efficiency variables, it is impossible (even in theory) to perfectly offset incentives for risk selection without mitigating incentives for efficiency at all. In practice, the regulator can never perfectly offset incentives for risk selection due to information asymmetries (see Chapter 2). Therefore, second best approaches need to be developed. In the US, the risk-adjustment formula that is applied by the Centers for Medicare and Medicaid Services (CMS) to pay Medicare Advantage plans is calibrated using only individuals in traditional (fee-for-service based) Medicare [Pope et al. (2011)]. However, this approach is only valid as long as the steepness of risk profiles in the two sectors is not too different. Two possible approaches to amend these problems were discussed by Van Kleef et al. (2010). The first approach entails calculating risk adjustment payments within plan types. This is possible only if the number of individuals per risk group is large enough within all plan types. The second approach is to augment observed expenditure in cost sharing plans by estimated 'true' cost savings. This approach requires the availability of an

universally trusted method of estimating 'true' cost savings, even for plans with small numbers of enrolles. The conditions for both approaches are currently not fulfilled in Switzerland. In order to provide practical guidance on the effective design of risk adjustment schemes in the case of contract differentiation, further research is needed.

Bibliography

Ai, C. and E. Norton (2000). Standard errors for the retransformation problem with heteroscedasticity. *Journal of Health Economics 19*, 697–718.

Ai, C. and E. Norton (2003). Interaction terms in logit and probit models. *Economics Letters 80*, 123–129.

Andersson, K., G. Bergstroma, M. Petzold, and A. Carlsten (2007). Impact of a generic substitution reform on patients' and society's expenditure for pharmaceuticals. *Health Policy 81*, 376–384.

Arrow, K. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review 53*, 941–973.

BAG (2009). Erlaeuternder Bericht zur Ordentlichen Revision des Bundesgesetzes über Arzneimittel und Medizinprodukte, HMG (Special report on the law on pharmaceuticals and medical devices). Technical report, http://www.bag.admin.ch/themen/medizin/.

Baltagi, B. and P. Wu (1999). Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory 15*, 814–823.

Beck, K. (1998). Risikoausgleich - Basis für einen sinnvollen Wettbewerb in der Krankenversicherung (Risk adjustment - Fondation of efficiency-enhancing competition in health insurance). Unpublished Manuskript.

Beck, K. (2004). *Risiko Krankenversicherung. Risikomanagement in einem regulierten Krankenversicherungsmarkt (The Risky Business of Health Insur-*

ance. Risk Management in a Regulated Market for Health Insurance). Bern: Haupt.

Beck, K., U. Kaeser, M. Trottmann, and S. von Rotz (2009). Efficiency gains thanks to managed care? Evidence from Switzerland. In B. Rosen, A. Israeli, and S. Shortell (Eds.), *Improving Health and Health Care. Who is responsible? Who is accountable? Conference Book of the 4th Jerusalem Conference on Health Policy*, pp. 539–557. Jerusalem: The Israel National Institute For Health Policy Research.

Beck, K., U. Kunze, and W. Oggier (2004). Selbstdispensation: Kosten treibender oder Kosten daempfender Faktor? (Drug dispensing by physicians: Cost increasing or cost decreasing?). *Managed Care 6*, 5–8.

Beck, K., S. Spycher, A. Holly, and L. Gardiol (2003). Risk adjustment in Switzerland. *Health Policy 65*, 63–74.

Beck, K., M. Trottmann, U. Käser, B. Keller, S. von Rotz, and P. Zweifel (2006). *Nachhaltige Gestaltung des Risikoausgleichs in der Schweizer Krankenversicherung (Sustainable Design of Risk Adjustment in Swiss Health Insurance)*. Bern: h.e.p. Verlag.

Beck, K. and P. Zweifel (1998). Cream-skimming in deregulated social health insurance: Evidence from Switzerland. In P. Zweifel (Ed.), *Health, The Medical Profession, and Regulation*, pp. 211–227. Dordrecht: Kluwer.

Ben-Akiva, M. and S. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press Series.

Blough, D., C. Madden, and M. Hornbrook (1999). Modeling risk using generalized linear models. *Journal of Health Economics 18*, 153–171.

Bodenheimer, T. (2006). Primary Care - Will it survive? *New England Journal of Medicine 355*, 861–864.

Bradley, D. and R. Lesu (2006). Optimal regulation of health systems with induced demand and ex post moral hazard. *Annales D'Economie et de Statistique 83*, 279–293.

Busato, A., P. Matter, B. Kunzi, and D. Goodman (2010). Supply sensitive services in Swiss ambulatory care: An analysis of basic health insurance records for 2003-2007. *BMC Health Services Research 10*, doi:10.1186/1472–6963–10–315,.

Cameron, A. and P. Trivedi (2005). *Microeconometrics* (1 ed.). New York: Cambridge University Press.

Cameron, A., P. Trivedi, F. Milne, and J. Piggott (1988). A microeconometric model of the demand for health care and health insurance in Australia. *Review of Economic Studies 55*, 85–106.

Cardon, J. and I. Hendel (2001). Asymmetric information in health insurance: Evidence from the National medical expenditure survey. *RAND Journal of Economics 32*, 408–427.

Chiappori, P., F. Durand, and P. Geoffard (1998). Moral hazard and the demand for physician services. *European Economic Review 42*, 499–511.

CMS (2004). Safe and effective approaches to lowering state prescription drug costs: Best practices among state medicaid drug programs. Technical report, Centers for Medicare & Medicaid Services.

Coscelli, A. (1998). Are market shares in drug markets affected by doctors and patients preferences for brands? *The Journal of Industrial Economics 28*, 349–369.

De Jaegher, K. and M. Jegers (2000). A model of physician behaviour with demand inducement. *Journal of Health Ecocomics 19*, 231–258.

Deb, P., M. Munkin, and P. Trivedi (2006). Bayesian analysis of the two-part model with endogeneity: Application to health care expenditure. *Journal of Applied Econometrics 21*, 1081–1099.

Deb, P. and P. Trivedi (2009). Provider networks and primary-care signups: Do they restrict the use of medical services? *Health Economics 18*, 1361–1380.

Decollogny, A. and M. Ruggli (2006). Die Situation der Generika in der Schweiz (Generic drugs in Switzerland). *Managed Care 3*, 23 – 26.

DeJaegher, K. and M. Jegers (2006). A model of physician behaviour with demand inducement. *Annales D'Economie et de Statistique 83*, 279–293.

Dowd, R., R. Feldman, S. Cassou, and M. Finch (1991). Health plan choice and utilization of health care services. *The Review of Economics and Statistics 73*, 85–93.

Duan, N., W. Manning, C. Morris, and J. Newhouse (1982). A comparision of alternative models for the demand for medical care. *The Rand Publication Series R-2754-HHS*.

Dummermuth, A. (1993). Selbstdispensation: Der Medikamentenverkauf durch Aerzte. Vergleiche und Auswirkungen unter besonderer Beruecksichtigung der Kantone Aargau und Luzern (Physician Dispensing: Comparisons and Consequences in Argovia and Lucerne). *Propharmacie, Cahiers de l'IDHEAP, No. 114*.

Eggleston, K. (2005). Multitasking and mixed systems for provider payment. *Journal of Health Economics 24*, 211–223.

Eichner, M. (1998). The demand for medical care: What people pay does matter. *AER Papers and Proceedings 88*, 117–121.

Ellis, R. (2008). Risk adjustment in health care markets: Concepts and applications. In M. Lu and E. Jonnson (Eds.), *Financing Health Care: New Ideas for a Changing Society*. Germany: Wiley-VCH Publishers.

Ellis, R. and T. McGuire (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics 5*, 129–151.

Ellis, R. and T. McGuire (1990). Optimal payment systems for health services. *Journal of Health Economics 7*, 135–151.

Ellis, R. and T. McGuire (1993). Supply-side and demand-side cost sharing in health care. *Journal of Economic Perspectives 9*, 375–396.

Emmerton, L., J. Marriott, T. Bessell, L. Nissen, and L. Dean (2005). Pharmacists and prescribing rights: Review of international developments. *Journal of Pharmacy and Pharaceutical Science 2*, 217–225.

Enthoven, A. (1978). Consumer-choice health plan. *The New England Journal of Medicine 12;13*, 650–658 and 709–720.

Evans, R. (1974). Supplier-induced demand: Some empirical evidence and implications. In M. Perlman (Ed.), *The Economics of Health and Medical Care*, pp. 162–173. Macimillan London.

Frondel, M. and C. Vance (2009). On marginal and interaction effects: The case of heckit and two-part models. *Ruhr Economic Papers*.

Gardiol, L., P. Geoffard, and C. Grandchamp (2005). Separating selection and incentive effects in health insurance. *PSE Working Papers 2005-38*.

Gardiol, L., P. Geoffard, and C. Grandchamp (2006). Selection and incentive effects: An econometric study of Swiss health-insurance claims data. In P. Chiappori and C. Gollier (Eds.), *Competitive Failures in Insurance Markets*, pp. 81–95. Cambridge, MA: MIT Press.

Garrett, J. M. (1997). Odds ratios and confidence intervals for logistic regression models with effect modification. *Stata Technical Bulletin 36*, 15–22.

Gerfin, M. and M. Schellhorn (2006). Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits - Swiss evidence. *Health Economics 15*, 1011–1020.

Goroll, A., R. Berenson, S. Schoenbaum, and L. Gardner (2007). Fundamental reform of payment for adult primary care: Comprehensive payment for comprehensive care. *Journal of General Internal Medicine 22*, 410–415.

Gourieroux, C., A. Monfort, E. Renault, and A. Trogon (1987). Generalized residuals. *Journal of Econometrics 34*, 5–32.

Griliches, Z. and I. Cockburn (1994). Generics and new goods in pharmaceutical price indexes. *The American Economic Review 84*, 1213 – 1232.

Guillaume, L. and R. Cooper (2008). Supplementary prescribing by community and primary care pharmacists: An analysis of PACT data, 2004 - 2006. *Journal of Clinical Pharmacy and Therapeutics 33*, 11–16.

Hardin, J. and J. Hilbe (2007). *Generalized Linear Models and Extensions* (2nd ed.). Stata Press.

Hay, J., R. Leu, and P. Rohrer (1987). Ordinary least squares and sample-selection models of health-care demand: Monte carlo comparision. *Journal of Business and Economic Statistics 5*, 499–506.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement*, pp. 120–137. New York: NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*, 153–161.

Hellerstein, J. (1998). The importance of the physician in the generic versus trade-name prescription decision. *The RAND Journal of Economics 29*, 108–136.

Hellstrom, J. and N. Rudholm (2010). Uncertainty in the generic versus brand name prescription. *Empirical Economics 38*, 503 – 521.

Holly, A., L. Gardiol, Y. Eggli, and T. Yalcin (2003). Health-based risk adjustment in Switzerland: An exploration using medical information from prior hospitalization. Institute of Health Economics and Management, Lausanne.

Hosmer, D. and S. Lemeshow (2000). *Applied Logistic Regression* (2 ed.). Wiley Series in Probability and Statistics.

Hunkeler, J. (2008). SL - Logistikmarge - Probleme und Reformansaetze im SD-Markt (Drug Margins for Physicians - Problems and Solutions). Technical report, Federal Office for Price Supervision.

Iizuka, T. (2007). Experts' agency problems: Evidence from the prescription drug market in Japan. *The RAND Journal of Economics 38*, 844–862.

Ikegami, N. and J. Cambell (1995). Medical care in Japan. *The New England Journal of Medicine 333*, 1295–1299.

IMS (2009). Swiss price indexes for molecules without patent protection. Technical report, www.interpharma.ch.

Iversen, T. and A. Ma (2010). Market conditions and general practitioners' referrals. *HERO Online Working Paper Series 8*.

Jensen, M. and W. Meckling (1992). Specific and general knowledge, and organizational structure. In L. Werin and H. Wijkander (Eds.), *Contract Economics*, pp. 251–274. Oxford: Blackwell.

Jones, A. (2000). Health econometrics. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics*, pp. 265–344. Amsterdam: Elsevier.

Kim, H., W. Chung, and S. Lee (2004). Lessons from Korea's pharmaceutical policy reform: The separation of medical institutions and pharmacies for outpatient care. *Health Policy 28*, 267–275.

Kim, H.-J. and J. P. Ruger (2008). Pharmaceutical reform in South Korea and the lessons it provides. *Health Affairs 4*, 260 – 269.

Kreier, R. and P. Zweifel (2010). Health insurance in Switzerland: A closer look at a system often offered as a model for the United States. *Hofstra Law Review 39*, 89–110.

Lamers, L. and R. Van Vliet (2003). Health-based risk adjustment: Improving the pharmacy-based cost group model to reduce gaming possibilities. *The European Journal of Health Economics 4*, 107–114.

Lee, L. (1978). Unionism and wage rates: A simultaneous model with qualitative and limited dependent variables. *International Economic Review 23*, 415–433.

Lehmann, H. and P. Zweifel (2004). Innovation and risk selection in deregulated social health insurance. *Journal of Health Economics 23*, 997–1012.

Leung, S. and S. Yu (1996). On the choice between sample selection and two-part models. *Journal of Econometrics 72*, 197–229.

Leutgeb, R., C. Mahler, G. Laux, A. Weschnetz, and J. Szecsenyi (2009). Health insurance discount contracts: Problems and risks for the general practitioner in the medical care of patients with chronic illness. *Deutsche Medizinische Wochenschrift 81*, 181–186.

Liu, Y., Y. Kao Yang, and C. Hsieh (2009). Financial incentives and physicians' prescription decisions on the choice between brand-name and generic drugs: Evidence from Taiwan. *Journal of Health Economics 28*, 341–349.

Lundin, D. (2000). Moral hazard in physician prescribing behavior. *Journal of Health Economics 19*, 639–662.

Manning, W. (1998). The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics 17*, 283–295.

Manning, W., N. Duan, and W. Rogers (1987). Monte carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics 35*, 59–82.

Manning, W. and J. Mullahy (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics 20*, 461–494.

Manning, W., J. Newhouse, N. Duan, E. Keeler, and A. Leibowitz (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review 77*, 251–277.

Marchand, M., M. Sato, and E. Schokkaert (2004). Prior health expenditures and risk sharing with insurers competing on quality. *Rand Journal of Economics 34*, 647–669.

Matsuda, R. (2008). Action plan for promoting generic substitution. *Health Policy Monitor Available at: www.hpm.org/survey/jp/a11/2*.

McGuire, T. (2000). Physician agency. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics, Ch.9*. North Holland.

McGuire, T. and M. Pauly (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics 10*, 385–410.

Morton-Jones, T. and M. Pringle (1993). Prescribing costs in dispensing practices. *British Medical Journal 306*, 1244–1246.

Newhouse, J. (1982). Is competition the answer? *Journal of Health Economics 1*, 109–116.

Newhouse, J. (2002). *Pricing the Priceless: A Health Care Conundrum*. Cambridge: MIT Press.

Newhouse, J., W. Manning, E. Keeler, and E. Sloss (1989). Adjusting capitation rates using objective health measures and prior utilization. *Health Care Financing Review 10*, 41–54.

Norton, E., H. Wang, and C. Ai (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal 4*, 154–167.

Nyman, J. (1999). The economics of moral hazard revisited. *Journal of Health Economics 18*, 811–824.

OECD (2006). Switzerland. OECD Reviews of Health Systems - www.oecd.org.

Olsen, R. (1980). A least squares correction for selectivity bias. *Econometrica 48*, 1815–1820.

Pauly, M. and B. Herring (1999). *Pooling Health Insurance Risks*. Washington D.C.: The AEI Press.

Pauly, M., A. Nickel, and H. Kunreuther (1998). Guaranteed renewability with group insurance. *Journal of Risk and Uncertainty 16*, 211–221.

Pope, G., R. Ellis, A. Ash, C. Liu, J. Ayanian, D. Bates, H. Burstin, L. Iezzoni, and M. Ingber (2000). Principal inpatient diagnostic cost group model for medicare risk adjustment. *Health Care Financing Review 21*, 93–118.

Pope, G., J. Kautter, M. Ingber, S. Freeman, R. Sekar, and C. Newhart (2011). Evaluation of the CMS-HCC risk adjustment model. *RTI International - CMS Contract No. HHSM-500-2005-00029ITO0006*.

Rabe-Hesketh, S., T. Toulopoulou, and R. M. Murray (2001). Multilevel modeling of cognitive function in schizophrenic patients and their first degree relatives. *Multivariate Behavioral Research 36:2*, 279–298.

Rischatsch, M. (2011). Drug dispensing by physicians - is there evidence of margin opimization? *Working Paper, Department of Economics, University of Zurich*.

Santesuisse (2009). Indices nach Laenderkorb: Im Durchschnitt tiefere Preise im Ausland (Cross-country indexes: lower foreign prices on average). Technical report, www.interpharma.ch.

Schellhorn, M. (2001). The effect of variable health insurance deductibles on the demand for physician visits. *Health Economics 10*, 441–456.

Schleiniger, R., T. Slembeck, and J. Bloechiger (2007). Bestimmung und Erklaerung der Kantonalen Mengenindizes der OKP-Leistungen (Cantonal price and quantity indices). ZHAW, School of Management.

Schokkaert, E. and C. Van de Voorde (2009). Direct versus indirect standardization in risk adjustment. *Journal of Health Economics 28*, 361–374.

Shen, Y. and R. Ellis (2002). How profitable is risk selection? A comparison of four risk adjustment models. *Health Economics 11*, 165–174.

Simon, C. and L. Blume (1994). *Mathematics for Economists*. New York: W.W. Norton & Company Inc.

Terza, J., A. Basu, and P. Rathouz (2008). Two-stage residual inclusion estimation: Addressing endogeneity in health econometric modeling. *Journal of Health Economics 27*, 531–543.

Van Barneveld, E., L. Lamers, R. Van Vliet, and W. Van de Ven (2000). Ignoring small predictable profits and losses: A new approach for measuring incentives for cream skimming. *Health Care Management Science 3*, 131–140.

Van de Ven, W., K. Beck, C. Van de Voorde, J. Wasem, and I. Zmora (2007). Risk adjustment and risk selection in Europe: Six years later. *Health Policy 83*, 162–179.

Van de Ven, W. and R. Ellis (2000). Risk adjustment in competitive health plan markets. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics*, pp. 755–845. Amsterdam: Elsevier.

Van Kleef, R., K. Beck, and F. Buchner (2010). Risk-type concentration and efficiency incentives: A challenge for the risk adjustment formula. *Geneva Papers on Risk and Insurance. Issues and Practice 35*, 503–520.

Van Kleef, R., K. Beck, W. Van de Ven, and R. Van Vliet (2008). Risk equalization and voluntary deductibles: A complex interaction. *Journal of Health Economics 27*, 427–443.

Van Kleef, R., W. Van de Ven, and R. Van Vliet (2006). Premium rebate in exchange for a voluntary deductible in social health insurance with risk equalisation: Community-rated of risk-rated? *Journal of Risk and Insurance 73*, 529–550.

Van Vliet, R. (1992). Predictability of individual health care expenditures. *Journal of Risk and Insurance 59*, 443–460.

Vatter, A. and C. Ruefli (2003). Do political factors matter for health care expenditure? A comparative study of Swiss cantons. *Journal of Public Policy 23*, 301–323.

Vella, F. (1993). A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review 34*, 441–457.

Vella, F. and M. Verbeek (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics 13*, 163–183.

Vella, F. and M. Verbeek (1999). Two-step estimation of panel data models with censored endogenous variables and selection bias. *Journal of Econometrics 90*, 239–263.

Welch, W. (1985). Regression toward the mean in medical care costs. Implications for biased selection in HMOs. *Medical Care 23*, 1234–1241.

Werblow, A., S. Felder, and P. Zweifel (2007). Population ageing and health care expenditures: A school of 'red herrings'? *Health Economics 16*, 1109–1126.

Winkelmann, R. (2004). Co-payments for prescription drugs and the demand for doctor visits - Evidence from a natural experiment. *Health Economics 13*, 1081–1089.

Wolfe, J. and J. Godderies (1991). Adverse selection, moral hazard, and wealth effects in the Medigap insurance market. *Journal of Health Economics 10*, 433–459.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge MA: MIT Press.

Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory 2*, 10–26.

Zweifel, P. (1985). Technology in ambulatory medical care: Cost increasing or cost saving? *Social Science and Medicine 21*, 1139–1151.

Zweifel, P. (1992). *Bonus Options in Health Insurance*. The Netherlands: Kluwer Academic Publishers.

Zweifel, P. (2007). The theory of social health insurance. *Foundations and Trends in Microeconomics 3*, 183–273.

Zweifel, P. and M. Breuer (2006). The case for risk-based premiums in public health insurance. *Health Economics, Policy and Law 1*, 171–188.

Zweifel, P., F. Breyer, and M. Kifmann (2009). *Health Economics*. Berlin: Springer.

Zweifel, P. and W. Manning (2000). Moral hazard and consumer incentives in health care. In A. Culyer and J. Newhouse (Eds.), *Handbook of Health Economics*, pp. 410–459. Amsterdam: Elsevier.

Zweifel, P. and M. Tai-Seale (2009). An economic analysis of payment for health care services: The United States and Switzerland compared. *International Journal of Health Care Finance and Economics 9*, 197–210.

Curriculum Vitae

Maria Trottmann was born on September 03, 1979 in Bad Toelz near Munich, Germany. From 2000 to 2005 she studied Economics at the University of Zurich and the University of Lausanne. After graduation, she worked as an econometrician for CSS Insurance, a large Swiss sickness fund. Her tasks included research projects in risk adjustment, prospective payment and managed care. Starting in 2006, she combined her job with part-time PhD studies in health economics at the University of Zurich (Adviser: Prof. Dr. Peter Zweifel). She also became a member of the European Risk Adjustment Network (RAN). In 2009, she left her position at CSS to become a visiting scholar at Boston University (Sponsor: Prof. Dr. Randall Ellis). In Boston, she collaborated with Verisk Health, a consulting firm that is focused on risk adjustment and provider payment. After returning to Switzerland, she continued to work for Verisk Health Germany, based in Munich. In October 2010, she left Verisk Health in order to finish her PhD studies. She presented her work at the RAN meetings in Berlin (2006), Jerusalem (2009), and Berlin (2010), the European Conference on Health Economics in Rome (2008), the International Jerusalem Conference on Health Policy in Jerusalem (2009), the Conference of the American Society of Health Economists in Ithaca, NY (2010), and at the World Congresses on Health Economics in Beijing (2009) and Toronto (2011).